International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

PROCESS MINING IN CANCER PREDICTION OF PROGNOSIS BY USING CLASSIFIERS BASED ON MIXED VARIABLES

S.Yamuna Rani[1],Dr.Sumagna Patnaik[2]
[1]*Assistant Professor, Government Degree college for Women, Gajwel, Telangana*
[2]*Professor, J.B.Institute of Engineering & Technology*

Abstract: A healthy decision-making process for health of the individual is a major challenge in this day and age of abundance of information all over the place. Machine learning, data mining and computation statistics have become among the top research areas which enable individuals empowered to make important decisions that will improve the results of any field. A high demand for data handling is evident in the healthcare sector, since the increase in the number of patients is proportional the increase in population and lifestyle changes. Strategies for early diagnosis and prognosis predictions of diseases are of great importance at the present time to offer better healthcare for the entire human population. Data mining is an advantage in the development of high-quality and effective models for applications to predict health. Since cancer has been everywhere in recent years, information from the cancer registry are being used as medical information in this research. The principal goal of the thesis is to create an efficient and effective classification model for predictions of prognosis for cancer. . The majority of the current system is based on diagnosis prediction model from surveys or screening data since the data is readily available and simple to gather due to the insensitivity of the variables involved in these studies. For prognosis prediction, it requires specific information about the patients who are receiving treatment for a recognized disease. Hospitals and community registries managed by the state constitute the most important sources for data gathering. Electronic hospital records that are well-maintained that include histopathology data are not accessible in India to researchers. Thus, data on cancer obtained from an US accessible data centre was utilized in this study to conduct all experiments.
This research study is a model that improves accuracy of predictions by utilizing the appropriate methods for data mining for each stage. Prognosis refers to the survival percentage of cancer patients generally, but is also a measure of how severe the cancer in the future timeline that the person. The two-fold goal of this study is to discover the key response variables that are a part of the prediction system used to determine the prognosis as well as enhance the predictive models. .
Keywords—Process Mining, Prognosis, Health care, Cancer Data.

1 INTRODUCTION:

Data mining is the process of obtaining patterns hidden in large amounts of data. It's a growing field that makes use of statistical visualization, machine learning as well as other data manipulation and knowledge extraction methods aimed to gain a better understanding of the patterns and relationships in the data [10]. The amount of data continues to increase in the present mobile and internet age, analytics is the buzzword of today's market, both in the IT as well as non-IT industries. Machine learning and data mining algorithms offer support in finding a solutions to the issues of massive data, and the majority of the applications are moving toward the healthcare field [4]. A high demand for data handling is present in the healthcare sector, since the increase in the number of patients is proportional the growth rate of the population and lifestyle changes.
An accurate model of the actual behaviour of a process increases the ability of defining and implementing the required process specifications in the HIS to support the process, defining any additional requirements that aren't already in the system and assisting in an analysis of the processes. Additionally, the author in [6] mentions that it is possible to extend the analysis by using different methods like organization mining, the automatic creation of simulation models as well as model extensions or repair of models, forecasting the behavior of a process, and making suggestions based on historical data.

DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

Healthcare is an area that is complex in its models that can be subject to substantial variation in time [2]. The healthcare industry is a complex area that has significant variations over time. These variations are caused by many factors, such as the diverse conditions of patients and the many methods and ways of actions that are completed by the healthcare resources (physician or nurse, as well as various healthcare specialists)

The capability of using methods to discover processes models and then analyzing their performance gives you opportunities to benefit from information contained in health care

In the world of medicine, cancer is the second leading causes of mortality, followed just by the heart condition. The mortality rates are higher than 50% of the rate of African as well as Asian countries, however, the prevalence is higher in Western countries, even however the mortality rate is lower. In the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute is an reliable source of information on cancer across the United States [8]. It is a cancer registry based on population which covers approximately 26 percent of the US population in a variety of geographic regions . It is the most comprehensive publicly accessible domestic cancer data collection. Breast cancer among women, lung cancer and prostate cancers in men are the most common cancers that impact the populace. Different prediction models based on SEER data sets have been developed in the past research studies.

The following factors have been the main motivations for conducting research into medical and cancer data.

* Exponential growth in patients during the past few years.
* The priority is to lower the risk of dying through predictive analysis.

CHALLENGES IN MEDICAL DATA MINING

In the medical world there is a constant demand for high precision, and in the latest studies, the expected precision has been closely accomplished through the use of data mining methods [6, 7]. Carcinogenesis is the scientists with a significant challenge and offers limited tools for controlling it. The cancer registries of India and around the world provide information and classification on all cancer patients so that they can provide data about the incidence of cancer, and give a framework to assess and controlling the effects and the extent of cancers in the community. SEER is the SEER data set was used in this study since the Indian cancer registry data set is not available to public research. The research options offered by traditional methods offered of these registries fail to reveal the fundamental characteristics of the data available. Data mining algorithms generally assist us uncover new relationships between data and discover patterns that can identify the conditions and symptoms that affect how to treat them and also prevent deaths.

This research is aimed at identifying and defining the instances in which process mining has been used in the cancer healthcare domain in order to provide an understanding of current state of the art in this field, aiding researchers to determine the best method to use in applying process mining techniques algorithms, methods, and tools, and highlighting the benefits that this method offers.

II RELATED WORK

2.1. Process Discovery Algorithms

There have been only a handful of reviews on processing mining and data mining in the field of healthcare. We have identified a few reviews that discuss the applications of data mining across a variety of medical fields [17-23]. In the context of process mining in health medical care, there's only an incredibly brief and specific research review of research studies that deal with the clinical pathway [24, 25]. But, there is no complete study that collects the details, characterizations and contexts of every case study in which process mining has been used within the healthcare field.

Process mining refers to a set of techniques that aims to gather useful insights from the data processes produce while they are being carried out. It acts as an interface between the field of process science (which encompasses areas like business processes management and operations research) in addition

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

to data science (which encompasses fields like predictive analytics and data mining) and results in methods for analyzing processes with data [88. Process mining is not a domain-specific approach, i.e, process mining techniques can be utilized to any field in which processes are used and the associated data are available. Healthcare, which is the topic of this article is a particular area where the application for process mining has been increasing.

Process mining in healthcare within the larger context. Process mining is about processes that can be described by a process model, like a model that represents the steps of the process, and the many routes a process may follow [22,23]. The sequence of actions in the process can be visualized in a variety of ways, e.g., using flowcharts [23] and Business Process Modeling notation (BPMN)

A significant effort can be seen in the scientifical production that is linked to Health Technology Assessment for healthcare management. There is lots to be researched in regards to the latest healthcare methods and this requires the development of appropriate processes and methodologies that are adapted to each country's context and the particular circumstances of each. This means developing and implementing innovative Health Technology Assessment techniques to assist in making the decision-making process in healthcare [2,5].

A few authors point out the difficulty in integrating the findings from Health Technology Assessment studies undertaken in different countries due to the possibility of reproducibility and replicability i.e. there are variations in the efficacy of different options, costs as well as the utilization of resources in the healthcare system, and epidemiological problems, and others. To address this issue The authors recommend a series of studies based on clinical data as well as data on the location's use of the region in question [6-8].

III. DATASET & PREPROCESSING METHODS

In this part we will look at the workflow for mining process and data set. Clinical trials that are randomized have become widely accepted and utilized for Health Technology Assessment. They've been long regarded as the best method to gives scientific evidence, and is as the "gold benchmark" to be used in Health Technology Assessment. But one of their major advantages, and in fact an issue is that they are used to specific populations in the targeted environment, which can differ from each country's specific socio-clinical or medical reality. This means that they're not able to show conclusive evidence about the effectiveness of a certain healthcare technologies particularly due to the sheer number of patients with comorbidities or demographics often differ from the characteristics of those utilized in clinical trials that use randomized control.

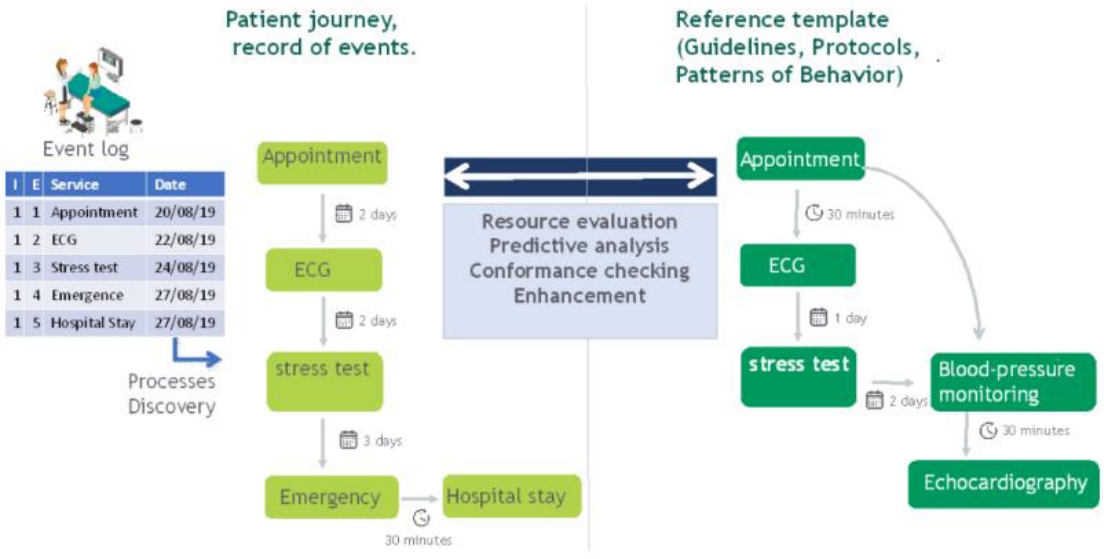

Fig. 1 Example of using Process Mining in Healthcare

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

3.1 DATA SET DESCRIPTION

The collection of cancer data for research has begun by various sources from India. In the Cancer Prevention and Early Detection (CPED) statistics and facts provided by the American Cancer Society provided a detailed information on the present state and the requirements for cancer research.

This SEER dataset provided by NCI has been extensively utilized in research studies for various classification issues. The particulars of this set of data were discussed, and an extensive pre-processing process has been utilized prior to the data being used for a classification study. The permission letter as well as the mail copy are included as an appendix for evidence of SEER.

The two other data sets used for reducing dimensionality were can be obtained from UCI Machine Learning Repository [19]. The details of the sets of data that were used is provided in the following section. Because these are considered to be the standard data that are used for classification and grading of the data, they have been directly used to test the problem of reducing the dimensionality by using NN PCA based PCA methods.

SEER Data Dictionary

The SEER data base used in this thesis focuses on the directory information of SEER_1973_2007_TEXTDATA. The research was started in 2009, and the directory that is based on period 1973-2021 research data was downloaded. The information in this directory is in ASCII text format, which is used for both incident-based and population data.

The file for the data dictionary named "SEERDIC.PDF" gave the layout of the text to aid in greater understanding. The files that include specific information about the site were used to further study the site in this thesis but leaving out the race and region specific information. The breast cancer sites colon and rectum cancers, stomach cancer Leukemia in the respiratory system, bladder cancer, and others were investigated in the beginning phase.

Table 1: Arrhythmia data set

Data Set Particulars	Specifications
Variable Type	Nominal, Real, Integer
# Instances	452
# Variable	280
# Missing Values	Yes
# Class labels	Class 01 : normal ECG Class 02 to 15 : arrhythmia types Class 16 : unclassified

3.2 EXPERIMENTAL SETUP

For the first phase of this research, MATLAB has been used to preprocess data and building base classifiers and for some improved classification methods, Rapidminer operators have been utilized. A GUI tool built into MATLAB was developed to assist in processing a large amount of preprocessing and feature extraction. The results are stored in Excel file format as well as in matrix information format within MATLAB to make it easier to perform interactions between the various tools.

The data samples range between 500 and 30000, with different seeds and scaling have been utilized for the implementation. The effectiveness of each model has been evaluated within the boundary of this region and then interpreted the results in the relevant chapters. The classifiers that are the base and enhanced classifiers for the SEER data sets were evaluated using prominent labels that are a part of all the features described in Table 2.

Table 2: Implementation features

Sample size	Classifiers	Data Set	Class Labels	Sampling
500		Breast	Survival	

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

1000	Naïve Bayes Decision Tree KNN	Cancer Colorectal Cancer Respiratory Cancer Mixed Type	Age at Diagnosis Multiple Primaries Stage Grade	Random Stratified Balanced
2000				
5000				
8000 -10000				
13000 - 16000				
20000				
30000				

Operational Flow of Classification Models

The operation sequence of every model that is discussed in this thesis follows the steps in Figure 2. Models differ in step 4 depending on the classifier type and the architectures they are based on.

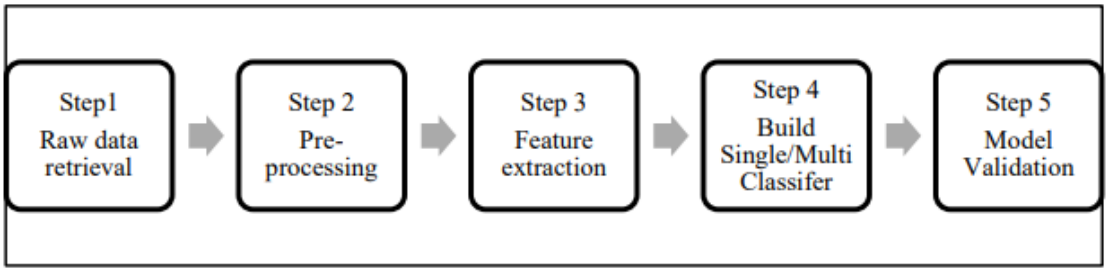

Fig 2: Operational flow of classification models

Validation Setup for Classification Models

The traditional evaluation methods were utilized in this study to testing the model's validity. Two different validation techniques were employed to test the effectiveness of the developed classifiers. For the basic classifiers enhanced classifiers, multi-label classification, the training and test splits of 60:40 has been adjusted, while 10 cross fold verification is used for the ensemble method [2,1]. The model's complexity is based on the most accurate and the lowest accuracy on the validation or test data set. Under fitting samples reduce the complexity while over fitting could make a model more complex. model.

The full evaluation procedure is described in the appropriate chapter of each model. The evaluation of performance of all classifiers has been illustrated using a confusion matrix that shows the amount of true Negatives (TN) as well as False Positives (FP) and False Negatives (FN) and True Positives (TP).

3.3 DATA PRE-PROCESSING

This section mainly discusses the extensive pre-processing phases of SEER data set.

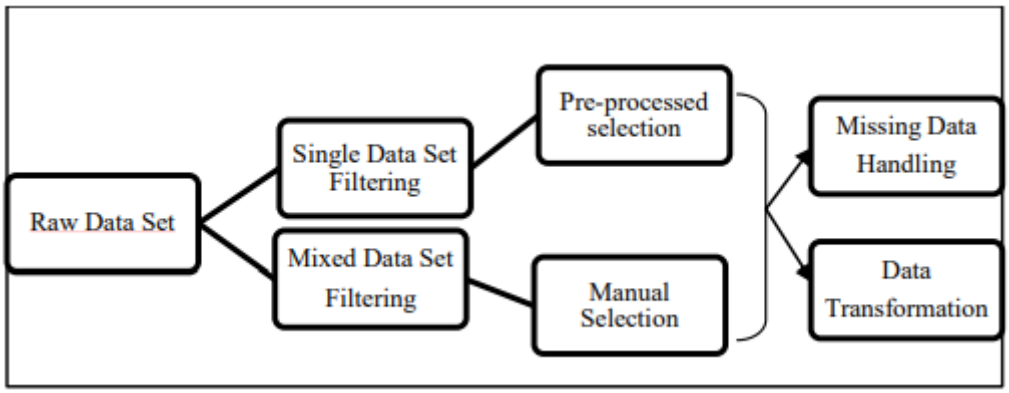

Fig 4 : Proprepocessing Procedure

The profile of the patient who was diagnosed with an incident is a single file that has 254 characters, which is the totality of 118 traits. The remaining race and region-based attributes are out of the scope

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

of this thesis. These 118 characteristics are numeral and nominal in their nature. Each of the 118 variables has been employed as features for the first phase. The attributes were diminished from 118 just 37 after doing a thorough processing of information.

Missing Data Handling

Apart from removing missing information through the formatting process and filters, two levels of missing have been dealt with by this part. In the course of SEER data update, some variables belonging to the group Extent of Disease (EOD) have been updated with Collaborative Stage (CS) classification since 2004. The variables have been combined to ensure that there is no missing or null data. The missing data for each attribute was handled using the method of mean imputation that is available in the Rapidminer tool. The processing of missing data and transformation reduced the number of variables in the previous step to just 62 variables.

FEATURE EXTRACTION FOR MIXED VARIABLES

The feature extraction process is a crucial step in any data mining issue. It is highly unlikely that all the variables are completely independent, and there is no structure of correlation among them. It is the responsibility of the analyst of data to protect against multi-collinearity. It is that occurs when some of the predictor variables are linked to one another. Multi-collinearity can cause instability within the solution space, creating incoherent results. Two well-known methods for feature extraction were used in the sections below. It is important to take into consideration the nature of variables that are used when performing the process of feature extraction.

Correlation Based Techniques

In the beginning of correlation-based feature selection process has been carried out using the output from earlier stage files [18, 19]. The attributes of the SEER data set have been re-dissected in the SEER group to provide greater analytical efficiency. Variables that aren't removed due to manual removal are discovered using the concept of correlation. The attributes were removed using the method of eliminating the most correlation variables one at a time. The equivalent operator in Rapidminer is applied to this process. An upper threshold value of 0.9 is set and any attributes that have a correlation greater than the threshold is eligible for elimination. Figure 4 illustrates the top 12 attributes ranked according to correlation. It was discovered that certain diagnostics variables from the same tests are duplicated within this data set. This results in reducing the size of the attribute to 44.

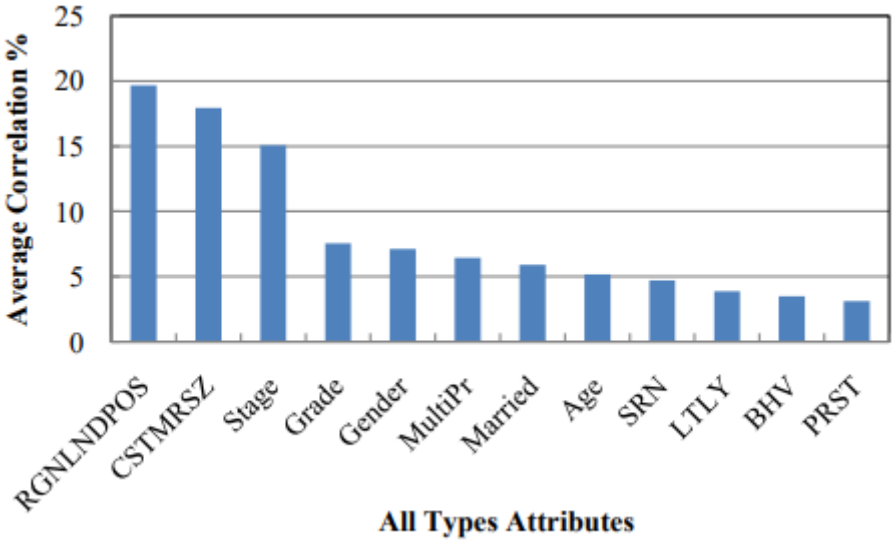

Fig 5 : Top average correlation score of all cancer attributes

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

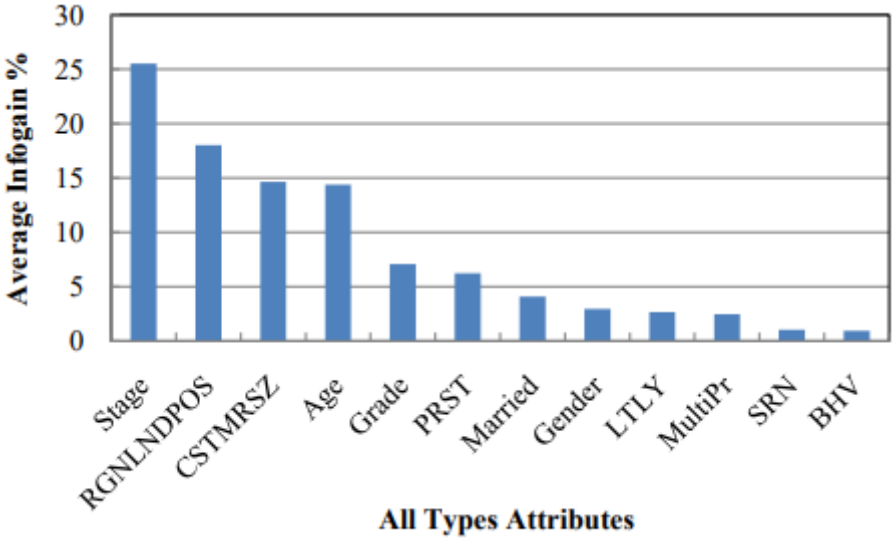

Fig 6 : Top average infogain score of all cancer attributes

IV EXPERIMENT AND RESULTS

The primary goal in this section is to determine the class names that are prominent and can help predict the prognosis of cancer among patients with positive diagnosis indicators.
Impurity Function: A function f can be considered to be an impurity function when it is defined using annum that comprise attributes (a1, A2, ..., and) when ai > zero, a >, ..., n, sum(aj)=1 so that f can be maximum only at (1/n, 1/n, ,..., 1/n) The minimum value of f is at (1 0, 0, ..., 0), (0, 1, 0, ..., 0), ..., (0, 0, ..., 0, 1) and the symmetry of f is (a1, a2 ..., the).

Posterior Probability: It's exactly proportional to both the the likelihood function and the probability prior to it. It is determined using the Bayes theorem that is described in equation

$$P\left(C_j/_X\right) = P(C_j)P\left(X/_{C_j}\right)/P(X)$$

The Distance Metric is used to calculate the distance between parameters of two samples, a variety of distance metrics have been utilized throughout the world of literature. For instance, the equation describes the Euclidean distance measurement for the attributes n. It determines the differences between one sample to the other. Another type of metric called similarity measure measures the similarity of two samples. It has been utilized to determine categorical variables in the equation.

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

$$sim(a,b) = \sum_{i=0}^{n} w_i S(a_i, b_i)$$

Where w_i is the weight factor and S is set as 1 if a_i=b_i and 0 otherwise

Algorithm:
Algorithm : Decision Tree Classification-DT(D,A,t)
Input : Training data set *D* = {(x1,c1),(x2,c2),.....(x3,c3)} and attribute list A= (a1, a2,..., ak)
Output : Decision Tree
Step 1 : Let t={}
Step 2 : Calculate A' the attribute with maximum info gain with D
Step 3 : Add A' to t

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

Step 4 : For all ai
in A repeat step 5-8
Step 5 : Branch at A' in t for all values of A'
Step 6 : Create subsets of Ds with values of a in A'
Step 7 : If D = ∅, then create leaf node with label c in C found in A'
Step 8 : Else call DT(Ds, A\ {A'}, t).

The variable t that is used in step 1 represents an attribute set which continues to add attributes during every iteration. A" represents the attributes list that has maximum information gain. Ds is the split list that is passed on to the next iteration.

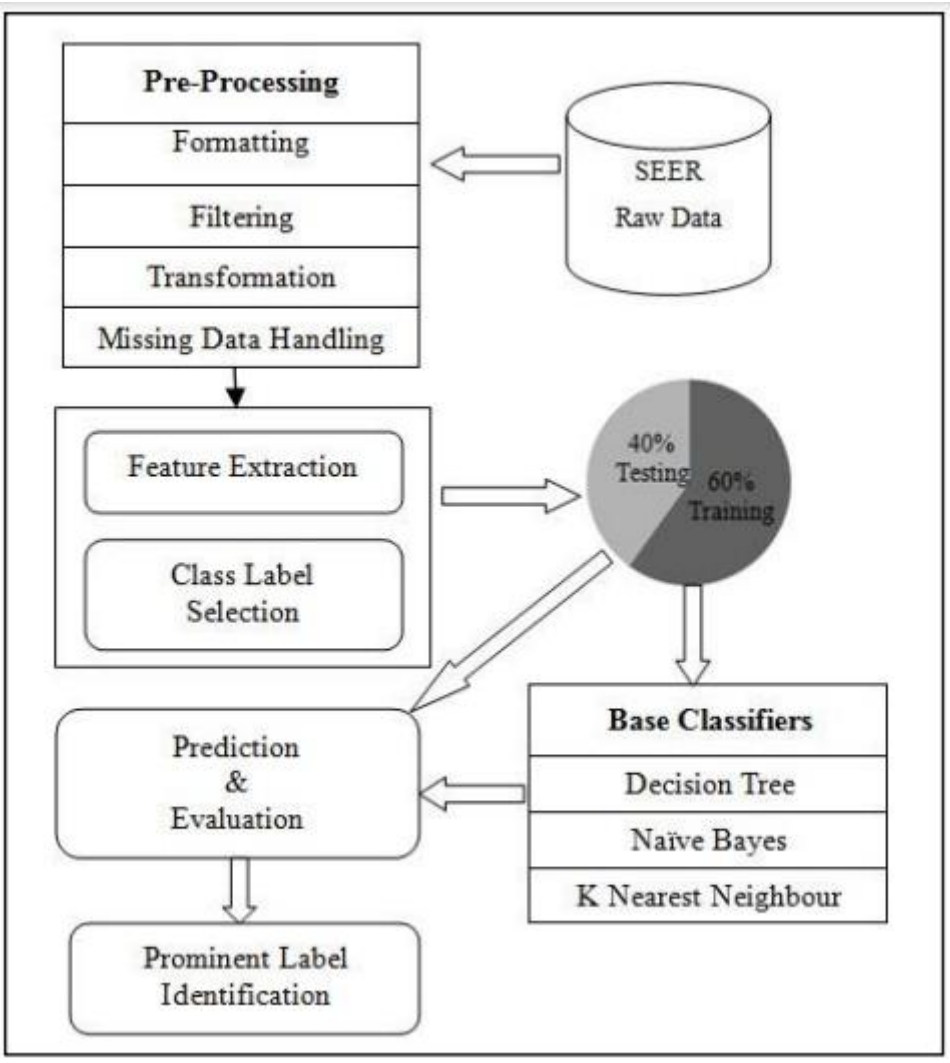

Fig 6 : Model architecture

The model is based on 60:40 ratios to testing as well as training examples. The top label with the highest fitness, which are chosen from the model will be regarded as the most prominent labels. The classification part of this model is developed using the Rapidminer to analyze any combination of cancer-related data sets derived from processed data.

Model suggested for selecting the most prominent cancer label prognosis prediction. The ten labels were initially selected as inputs to the model, and it was kept an average accuracy threshold at 50% for prominent labels. This means that just the five top labels were classified as prominent labels. The label's grade was included in this report, but the label grade has been disregarded for further study because the performance of the label has fluctuated depending on the size of the sample and the

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

selection of the classifier. Due to the fact that it is inconsistent the label grade was removed in the future.

Table 3: Top five prominent class labels

Legend	Class Label	No of Classes
1	Survival	3
2	Age at Diagnosis	5
3	Multiple Primaries	5
4	Stage	5
5	Grade3	5

The above three results from several of the results of the experiment was made into an illustration of each collection. The first result represents the most effective classifier for any combination, the next result of the frequent data sets that are used in all classifiers , and the third one is the most prominent name in every classifier.

Three possible combinations of the results are presented in different ways of visualization. Line graphs show the progression of the results, while the bar scale illustrates the intergroup comparison between classifiers while the tabular view illustrates real data elements.

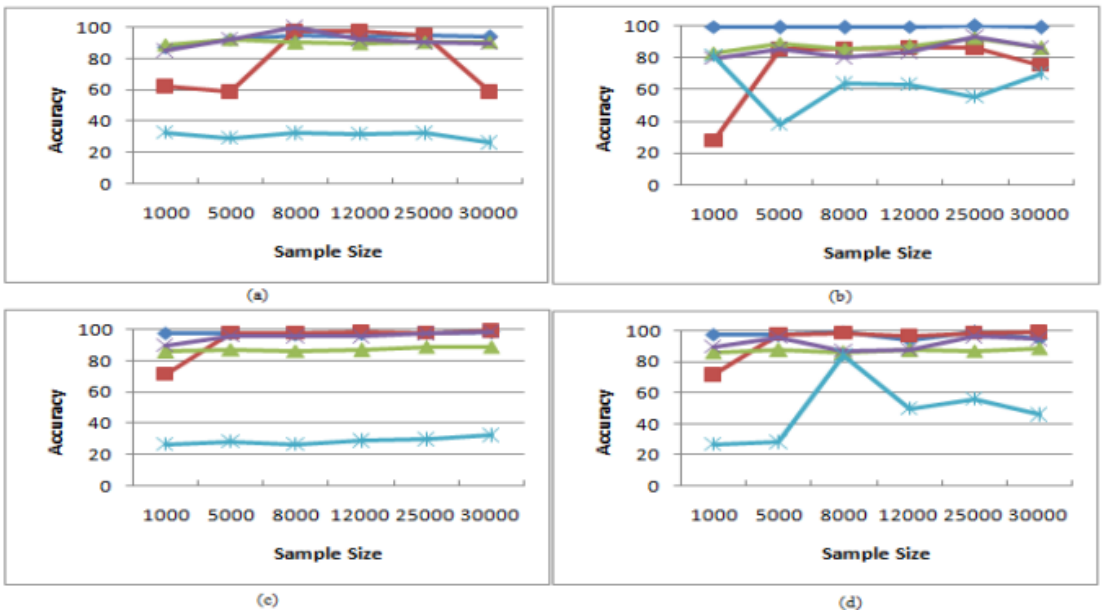

Fig 7 : Comparisons of all prominent labels in the following order of data set (a) Breast Cancer (b) Colorectal Cancer (c) Respiratory Cancer (d) Mixed Cancer (e) Legend numbers as given in Table 4.

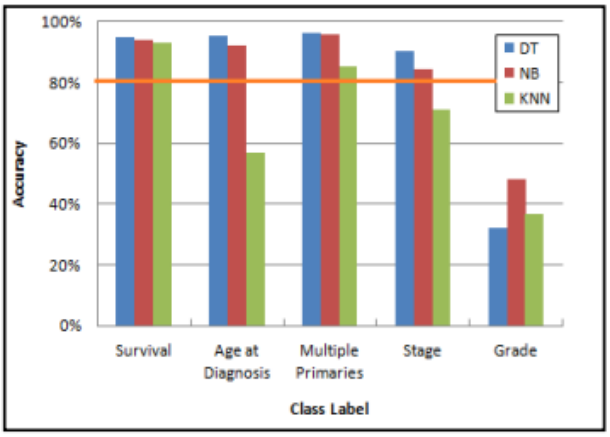

Fig 8 : Comparative performance of prominent labels using all classifiers for breast cancer data set with sample size 25000

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

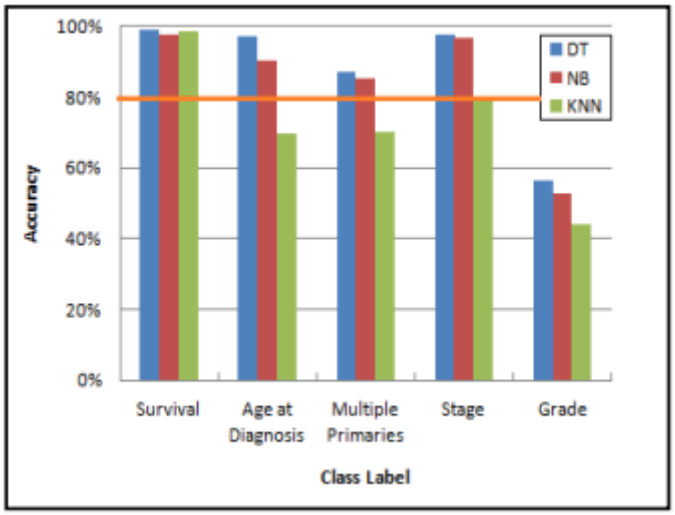

Fig 9 : Comparative performance of prominent labels using all classifiers for mixed cancer data set with sample size 25000

The efficiency of decision trees as well as Naive Bayes classifiers is superior to the k-NN classifier shown in Figure 8 and Figure 9. The horizontal line with 80percent accuracy is the threshold at which it determines the label labels for class as the most prominent ones to further research. Based on Figure 8 as well as Figure 9,, it is clear that the grade of the label is not a reliable indicator to determine the prognosis for the individual as well as in the various cancer kinds. Therefore, the grade was removed from the top five, and just the four top categories have been determined to be the most prominent.

Table 4: Survival prediction accuracy of all cancer types

Classifier	Breast	Colorectal	Respiratory	Mixed	Sample Size
DT	95.14	99.34	97.21	**99.78**	10000
NB	93.23	96.43	95.12	**99.34**	
Knn	92.63	99.08	95.21	**99.65**	
DT	93.45	**99.89**	97.43	94.23	20000
NB	93.45	96.98	97.67	95.34	
Knn	93	**99.76**	**99.99**	92.34	

The results of the analysis in Figure 7 shows that the survival of the label scores the top position of the four labels, therefore the performance data of this label across every data set and the three classifiers have been listed in Table 4. The information above was presented because of the large variances in outcomes found across every data set in samples that fall between 10000 to 20000. The bold text in Table 4 highlights best cancer types in each classifier . The underlined text highlights the top classifier for each type of cancer.

V CONCLUSION

The aim of identifying important class labels for predicting the prognosis for disease patients was achieved by with the well-known classifier used in data mining. The four most prominent labels that were that were filtered out of ten labels indicate the achievement of the initial stage of research and provides the possibility to further explore SEER data. SEER data set since the survival rate is only one of those considered by numerous research researchers throughout the world. Based on the accuracy average of the whole study the survival rate and multiple primaries, stage as well as age of diagnosis within the specified order were identified as the primary response variable, while the grade

DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

scored very low, and was eliminated. The efficacy of all four labels was evaluated using a simple random selection of samples that range from 1000-30,000. This means that the theory "More than one well-known label is used in the field of prognosis prediction" has been proven. The positive outcome of this research enables practitioners and tools of software to select the appropriate responses variables from the pool of prominent labels used for prognosis prediction. The main limitation of this model is that it does not make it easier to standardize the entire process general factors, and for a variety of ailments.

VIII REFERENCES

[1] W. M. P. van der Aalst, Process mining: Discovery, conformance and enhancement of business processes. Springer Science & Business Media, 2011.
[2] W. M. P. van der Aalst, A. J. M. M. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 9, pp. 1128–1142, 2004.
[3] C. W. Gunther and W. M. P. van der Aalst, "Fuzzy mining–adaptive ¨ process simplification based on multi-perspective metrics," in Business Process Management. Springer, 2007, pp. 328–343.
[4] J. M. E. M. Van der Werf, B. F. van Dongen, C. A. J. Hurkens, and A. Serebrenik, "Process discovery using integer linear programming," in Applications and Theory of Petri Nets. Springer, 2008, pp. 368–387.
[5] A. J. M. M. Weijters and J. T. S. Ribeiro, "Flexible heuristics miner (fhm)," in Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining. IEEE, 2011, pp. 310–317.
[6] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering block-structured process models from event logs - a constructive approach," in Application and Theory of Petri Nets and Concurrency, ser. LNCS. Springer, 2013, pp. 311–329.
[7] R. P. J. C. Bose and W. M. P. van der Aalst, "Abstractions in process mining: A taxonomy of patterns," in Business Process Management, ser. LNCS. Springer, 2009, pp. 159–175.
[8] C. W. Gunther, A. Rozinat, and W. M. P. van der Aalst, "Activity ¨ mining by global trace segmentation," in Business Process Management Workshops, ser. LNBIP. Springer, 2010, pp. 128–139.
[9] B. F. van Dongen and A. Adriansyah, "Process mining: Fuzzy clustering and performance visualization," in Business Process Management Workshops, ser. LNBIP. Springer, 2010, pp. 158–169.
[10] C. W. Gunther and H. M. W. Verbeek, "XES-standard definition," ¨ BPMcenter.org, 2014.
[11] T. van Kasteren, A. Noulas, G. Englebienne, and B. Krose, "Accurate ¨ activity recognition in a home setting," in Proceedings of the 10th International Conference on Ubiquitous Computing. ACM, 2008, pp. 1–9.
[12] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in Pervasive Computing, ser. LNCS, A. Ferscha and F. Mattern, Eds. Springer, 2004, pp. 158– 175.
[13] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in Pervasive Computing, ser. LNCS, A. Ferscha and F. Mattern, Eds. Springer, 2004, pp. 1–17.
[14] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," ACM SIGKDD Explorations Newsletter, vol. 12, no. 2, pp. 74–82, 2011.
[15] R. Poppe, "A survey on vision-based human action recognition," Image and Vision Computing, vol. 28, no. 6, pp. 976–990, 2010.
[16] L. Chen and C. Nugent, "Ontology-based activity recognition in intelligent pervasive environments," International Journal of Web Information Systems, vol. 5, no. 4, pp. 410–430, 2009.
[17] D. Riboni and C. Bettini, "OWL 2 modeling and reasoning with complex human activities," Pervasive and Mobile Computing, vol. 7, no. 3, pp. 379–395, 2011.

DOI:10.46647/ijetms.2022.v06i05.138 ISSN: 2581-4621

[18] T. van Kasteren and B. Krose, "Bayesian activity recognition in ¨ residence for elders," in Proceedings of the 3rd IET International Conference on Intelligent Environments. IEEE, 2007, pp. 209–212.
[19] J. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in Proceedings of the 18th International Conference on Machine Learning. Morgan Kaufmann, 2001.
[20] L. R. Rabiner and B.-H. Juang, "An introduction to hidden Markov models," ASSP Magazine, vol. 3, no. 1, pp. 4–16, 1986.
[21] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," Machine Learning, vol. 29, no. 2-3, pp. 131–163, 1997.
[22] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," Pervasive Computing, vol. 9, no. 1, pp. 48–53, 2010.
[23] W. Reisig, Petri nets: an introduction. Springer Science & Business Media, 2012, vol. 4.
[24] T. Murata, "Petri nets: Properties, analysis and applications," Proceedings of the IEEE, vol. 77, no. 4, pp. 541–580, 1989.
[25] H. M. W. Verbeek, J. C. A. M. Buijs, B. F. Van Dongen, and W. M. P. van der Aalst, "ProM 6: The process mining toolkit," in Proceedings of the Business Process Management Demonstration Track, ser. CEURWS.org, 2010, pp. 34–39.