Hybrid Privacy Preserving Mechanism: An Approach to Protect Health Care Data

M. Rameshkumar[1] and V. Lakshmipraba[2]
[1]Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India
[2]Assistant Professor, Department of Computer Science, Rani Anna Government Arts College, Tirunelveli, Tamil Nadu, India

Abstract **- With a lot of clinical information produced regularly, efficient methods have to be used to unearth significant data. Securing the information from the unapproved clients is also a major task to be achieved. Though lot of research has been carried out in these areas separately, a Hybrid architecture which combines both the features – efficiency and security is not widely found. The proposed architecture has been built taking these aspects into consideration. A proficient strategy for cross breed information mining method is applied here which incorporates the combination of Navie Bayesian classifier and Homomorphic encryption calculation.**

Keywords: **Homomorphic, Bayesian classifier, Privacy preserving, Prediction of cancer**

I. INTRODUCTION

At present data-mining become the most valuable tools to extract and manipulate data and for patterns to produce useful information for decision-making. Nearly all areas of life activities demonstrate a similar pattern whether the activity is finance, banking, marketing, retail sales, production, population study, employment, human migration, health sector, monitoring of human or machines, science or education, all have ways to record known information but are handicapped by not having the right tools to use this known information to tackle the uncertainties of the future.

Breakthroughs in data-collection technology, such as bar-code scanners in commercial domains and sensors in scientific and industrial sectors, have led to the generation of huge amounts of data [1].

The knowledge and intelligence techniques can be achieved by mining large volume of data in databases. For example, NASA's Earth observing System, which is expected to return data at the rate of several gigabytes per hour by the end of the century, has now created new needs to put this volume of information to use in order to help people make better choices in that area [2]. These needs include the automatic summarization of data, the extraction of the "essence" of information stored, and the discovery of patterns in the raw data. These can be achieved through data analyses, which involve simple queries, simple string matching, or mechanisms for displaying data [3].

Estimates of future values of business variables are needed. The commodities industry needs prediction or forecasting of supply, sales, and demand for production planning, sales, marketing and financial decisions [4]. In a production or manufacturing environment, we battle with the issues of process optimization, job-shop scheduling, sequencing, cell organization, quality control, human factors, material requirements planning, and enterprise resource planning in lean environments, supply-chain management, and future-worth analysis of cost estimations, but the knowledge of data-mining tools that could reduce the common nightmares in these areas is not widely available. It is worthwhile at this stage to state that extracting the right information from a set of data using data-mining techniques is dependent not only on the techniques themselves but on the ingenuity of the analyst.

In the modern livings privacy-preserving data mining took major role in research, because of using large volume of sensitive information on the internet. Different kinds of number of algorithmic techniques are available for privacy-preserving data mining. Aggarwal C.C, et. al[8] provides a review of the traditional mechanism for privacy and methods for the concept of randomization, distributed privacy-preserving and *k*-anonymization, data mining. The computational and theoretical limits associated with privacy-preservation also illustrated.

J.B. Awotunde, *et. al.,* [9] provide one of the major problems that both the developed and under-developed countries are facing is the difficulties of treating ill health people. There is shortage of medical expertise in various hospitals, most of these countries are spending affluence of their resources to meet this challenge but still they are unable to meet the demands of providing good medical services for their people. It has become of great concern to find a lasting solution to the problem of traditional method of medical diagnostic which is characterized by inaccuracy and imprecision. This approach signifies the medical diagnostic system functions through fuzzy system; so as to enhance the accuracy and precision of medical diagnosis. Current technological based medical diagnosis systems can be created to help people's drugs prescription, patient's records maintenance in the medical sector.

S Das, *et. al.,* [10] described a online based medical diagnostic support system (MDSS) through this health care supports provided for people living in rural areas. This approach uses fuzzy and its novel approach for medical

diagnosis. Subsequently, based on the proposed approach a web-based MDSS is developed. The MDSS comprises of a knowledge base (KB) and intuitionistic fuzzy inference system (IFIS). The precisions and certainty of medical data can be observed based on the observation.

K.M. Al-Aidaroos, *et. al.*, [11] illustrated as Medical centers collect maximum size of voluminous amounts of electronic data with more complicated form. The specific characteristics of medical data are very challenging and attractive. The databases can be classified more effectively using using of Naive Bayes (NB) classification algorithms and has been successfully applied to many medical problems.

The empirical comparison made Naïve bayes algorithm with Logistic algorithm, Neural Network and simple rule based algorithm and decision tree algorithms on selected databases. NB algorithm produced the on specific dataset. This research motivate Naïve bayes classification suits can fulfill all medical database requirements. The unrealistics independence assumptions and different individual techniques can be resolved by hybriding Naïve bayes with other proposed algorithm.

Tung-Shou Chen, *et. al.,[*12] depicted a hybrid protection scheme to protect privacy information and knowledge of cluster data in mined data. The scheme illustrates integrating the privacy-preserving data mining technique along with the knowledge-preserving anti-data mining technique. Hierarchical clustering is used and the clustering structure is manipulated by perturbation to create original data where the mined data has the appearance of similar information and knowledge from the original dataset but with misleading and non-useful contents.

II. HYBRID PRIVACY PRESERVING PROPOSED APPROACH

Since the database contains all kinds of datasets it is important to extract the data according to user request. In this work, hospital management application has been focused, so the database contains all the record of each patients. At the same time, it is important to secure the database from unauthorized user. Taking these aspects into consideration, a hybrid data mining technique is proposed, which is the combination of Homomorphic algorithm and Naive Bayes classifier. These two algorithms were chosen because the functionality of both the algorithms is based on probability. Furthermore, Naive Bayes classifier finds wide application for classification of variety of datasets. The Homomorphic algorithm uses probability based encryption to generate cipher text and Naive-bayes algorithm calculate prior probability which makes the encryption very efficient and fast. As a hybrid approach, these two algorithms are blended together and the results are analyzed. In this work, the benchmark dataset with cancer data is considered. The work has been implemented in the following three stages:
Stage 1: Constructing FDT to classify the data as normal and abnormal based on dataset characteristics.
Stage 2: Data Classification using Naive Bayes Classification for effective data retrieval and Data Encryption using Homomorphic Algorithm.
Stage 3: Data transfer using windowing and sliding algorithm.

The principle goal of the proposed work is to secure the available medical information from the unapproved clients and arrange the information with a specific end goal to reduce time and to accomplish exactness. Fig. 1 exhibits the general design for the proposed work.

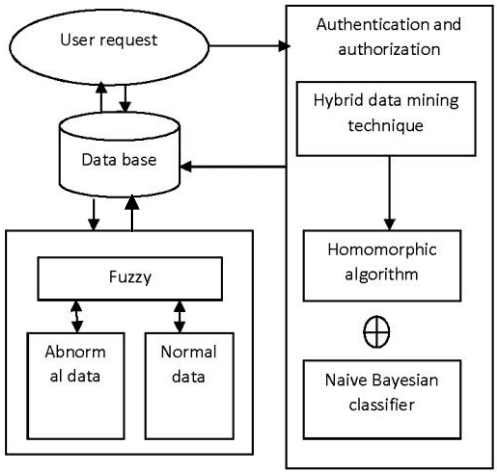

Fig.1. System architecture of the proposed system

Choosing the proper dataset plays a vital role in implementing the algorithm. The document to be chosen must be in such a way that it satisfies the following precondition:

1. Basic consistency check of the document to guarantee that field names and record structures are homogeneous between various source documents.
2. Distinguish odd information components, like missing qualities and wrong information
3. Inspect the information for pertinence and fulfilment. Checking for significance guarantees that appropriate data is incorporated to perform the study objective. For instance, information not containing any express data about the patient's kind of malignancy. For this situation, the information must be further handled to figure out whether a blend of various fields can infer and substitute the missing data. On the off chance that it is unrealistic, the dataset might be regarded wrong and be disposed of.
4. The information must be analyzed for subjective consistency. The goal is to guarantee the dataset overall is significant and helpful. For instance, when digging for the onset time of prostate malignancy, a sort of tumor showing up solely on male patients, subjective consistency may demonstrate that the information incorporates female patients.

Satisfying all the above requirements, a benchmark dataset has been taken into consideration which is described in the following section.

A. Dataset Description

A benchmark dataset with electrical impedance measurements in samples of freshly excised tissue from the breast is taken into consideration with 106 instances and 10 attributes with nine feature attributes and one class attribute. This data set has been studied by J. Jossinet and his team members for classification of breast tissues as normal and pathological. Six classes of freshly excised tissue were studied by the researchers using electrical impedance measurement which has been illustrated in Table I.

TABLE I CLASS ATTRIBUTES IN DATA SET

Class Name	Dataset class	No. of instances
Carcinoma	Car	21
Fibro-adenoma	Fad	15
Mastopathy	Mas	18
Glandular	Gla	16
Connective	Con	14
Adipose	Adi	22
Total instances		106

Impedance measurements were made at the frequencies: 15.625, 31.25, 62.5, 125, 250, 500, 1000KHz. These measurements were plotted by the researchers in the plane that constitute the impedance spectrum from where the features are computed. The various features to be extracted using the Bayesian classifiers are, listed in Table II.

TABLE II FEATURES IN DATASET

Features in dataset	Description
I0	Impedivity (ohm) at zero frequency
PA500	phase angle at 500 KHz
HFS	high-frequency slope of phase angle
DA	impedance distance between spectral ends
AREA	area under spectrum
A/DA	area normalized by DA
MAX IP	maximum of the spectrum
DR	distance between I0 and real part of the maximum frequency point
P	length of the spectral curve

The entire dataset with 106 instances is illustrated in Appendix A. For the considered dataset, Bayesian progressive model is applied to develop a classifier from a given arrangement of preparing case with class names. Bayesian Classifier applies the Bayes Theorem and prepared the occasions in a dataset where each and every occasion is distinguished by its characteristic set and the class variable is used to decide the particular characteristic set. The classification of cancer data has been illustrated in this work. The classified dataset is then encrypted using Homomorphic encryption algorithm for providing security and the encrypted data is stored in the database.

B. Naive Bayesian Classifier

The Naive Bayes calculation is a straightforward probabilistic classifier that ascertains an arrangement of probabilities by tallying the recurrence and blends of qualities in a given information set. The likelihood of a particular element in the information shows up as a part in the arrangement of probabilities and is determined by figuring the recurrence of every element esteem inside a class of a preparation information set. The preparation dataset is a subset, used to prepare a classifier calculation by utilizing known qualities to anticipate future, obscure qualities.

For the considered dataset, a factual displaying system, called the Hierarchical Association Rule Model (HARM) estimates a patient's conceivable future indications given the patient's information. The key of this method is a Bayesian progressive model. Bayesian systems (BN) on coordinated non-cyclic chart encodes a joint likelihood appropriation over an arrangement of properties X. The BN is utilized to develop a classifier from a given arrangement of preparing case with class names. Bayesian Classifier applies the Bayes Theorem to prepare the occasions in a dataset and arrange new occurrences to the most likely target esteem. Every occasion is distinguished by its characteristic set and a class variable. The objects are classified based on the user requirements which easily helps them to classify new cases as they arrive. This helps them to decide, to which class label they belong, based on the currently exiting objects. The classification of cancer data is illustrated as follows. Calculating the priors (i.e. the probability of the object among all objects) is based on the previous experience.

$$\text{Prior probability of cancer} = \frac{\text{number of cancer data}}{\text{total number of data}}$$

$$\text{Prior probability of other disease} = \frac{\text{number of other disease data}}{\text{total number of data}}$$

Having formulated prior probability a new object is classified. Since the objects are well clustered, To measure the likelihood between the given classified data. Then we calculate the number of points belonging to each class label.

$$\text{Likelihood(other disease)} = \frac{\text{number of other disease in vicinity of classified set}}{\text{total number other disease data}}$$

In the Bayesian analysis, the final classification is produced by combining both sources of information (i.e. the prior and the likelihood) to form a posterior probability using Bayes Rule.

Posterior probability of classified data being cancer = Prior probability of cancer ✕ Likelihood (cancer)
Posterior probability of classified data other disease = Prior probability of other disease✕ Likelihood (cancer)

In probability theory, Bayes theorem relates the conditional and marginal probabilities of two random events. It is often used to compute posterior probabilities given observations. Let x = (x_1 , x_2 , . . . , x_d) be the d-dimensional instance which has no class label, and our goal is to build a classifier to predict its unknown class label based on Bayes theorem. Let C = {C_1, C_2, . . . , C_K} be the set of the class labels.

$P(C_k)$ is the prior probability of C_k where (k = 1, 2, ..., K) are inferred before new evidence; $P(x|C_k)$ is the conditional probability of seeing the evidence x if the hypothesis C_k is true. The technique for constructing such classifiers to employ Bayes' theorem is as follows: $P(C_k|x) = \frac{P(x|Ck)P(Ck)}{P(x|Ck')P(Ck)}$

A naive Bayes classifier assumes that the value of a particular feature of a class is unrelated to the value of any other feature, so that $P(x|C_k) = \prod_{k=1}^{n} P(Xj|Ck)$ Probability density estimation constitutes an unsupervised method that attempts to model the underlying density function from which a given set of unlabeled data can be generated.

The characteristic set X considered in this work is,

X = { I0, PA500, HFS, DA, AREA, A/DA, MAX IP, DR, P} and
class variables are as follows:
{ Car, Fad, Mas, Gla, Con, Adi }
P(Car/I0), P(Fad/I0) ..... P(Adi/I0),
P(Car/PA500), P(Fad/ PA500)..... P(Adi/ PA500),
...
...
P(Car/P), P(Fad/P)..... P(Adi/P)

It must be figured for each of the class variable qualities in light of the data accessible in the preparation information. In the event that P(Car/I0) >= P(Fad/I0) >= P(Adi/I0), for six class values, then the new occasion is ordered to Car or Fad… or Adi likewise. This classifier appraises the class-contingent likelihood by expecting that the traits are restrictively free. Each of the qualities can be clear cut or numeric in nature. Fig.2 depicts the Naive bayes classification.

C. Homomorphic Algorithm

Homomorphic encryption is a form of encryption that allows computations to be carried out on ciphertext, thus generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext. Completely Homomorphic Encryption plans empower the calculation of important operations on encoded information without watching the real information. Illustrated in Fig. 3.

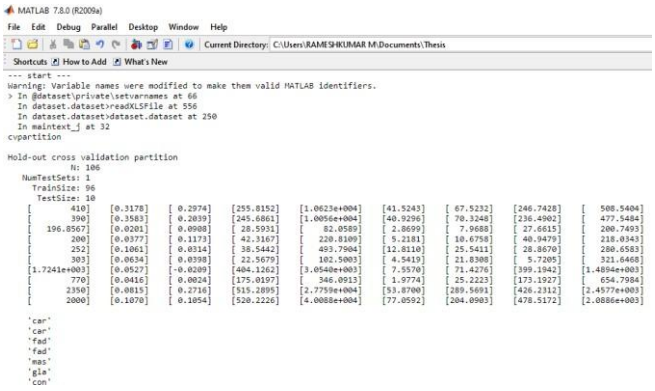

Fig. 2 Naive bayes classification

A multiplicative homomorphic cryptosystem has an encryption capacity E that fulfills the accompanying property:

$E(M_1) * E(M_2) = E(M_1*M_2)$ where M_1 and M_2 are plain instant messages.

Making utilization of homomorphic encryption will ensure security for patient's information. For safeguarding information mining they consider an extremely fundamental situation where there are n patients $(P_1...P_n)$ and every client Pi has a Boolean quality d_i . The pecialis might want to discover what number of di 's are 1's and what number of are 0's without uncovering any of the di 's.

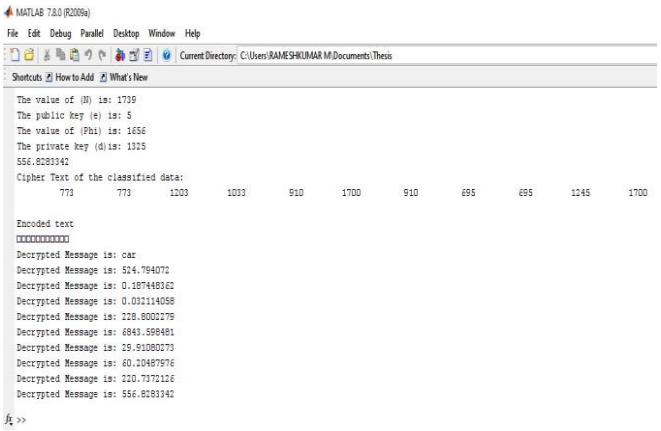

Fig.3 Encryption/decryption of user data

The hybrid privacy preserving mechanism employing Naïve-Bayes and Homomorphic encryption algorithms provides security by keeping the patient data in the encrypted form and also enhances execution. We give the report of the patient which demonstrates whether that specific patient having the abnormality or not. Impedance measurements that were made at the frequencies: 15.625, 31.25, 62.5, 125, 250, 500, 1000 KHz in the given dataset is analysed on its characteristics values. It demonstrates the likelihood of every information characteristic for the anticipated state. The grouping precision rates for the datasets were measured. For instance, in the order problem with two-classes, positive and negative, a solitary expectation has four conceivable outcomes. The True Positive rate (TP) and True Negative rate (TN) are precise

arrangements. A False Positive (FP) happens when the result is erroneously anticipated as positive when it is really negative. A False Negative (FN) happens when the result is inaccurately anticipated as negative when it is really positive.

1. Accuracy - It refers to the total number of records that are correctly classified by the classifier.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

2. Classification error - This refers to the misclassified datasets from the accurately classified records.
3. True Positive Rate (TP): It corresponds to the number of positive examples that have been accurately predicted by the classification model.
4. False Positive Rate (FP): It corresponds to the number of negative examples that have been inaccurately predicted by the classification model.
5. Precision - is the fraction of retrieved instances that are correlated.
   Precision= TP/TP+FP
6. Recall- is the fraction of correlated instances that are retrieved.

Recall = TP/TP + FN

Cancer disease prediction depicted in Table III and output measure shown in Fig. 4.

TABLE III PREDICTION

Prediction		Disease	
		+	-
Cancer Prediction Test	+	True Positive (TP)	False Positive (FP)
		False	True
	-	Negative (FN)	Negative (TN)

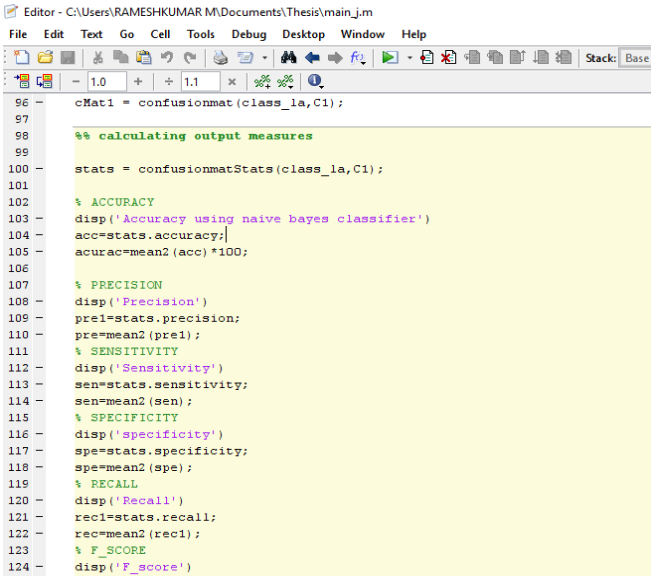

Fig. 4 Output measure calculation

III. PERFORMANCE ANALYSIS OF IMPLEMENTATION RESULTS

The comparison of different parameters of existing algorithms with hybrid algorithm is given in the Table IV and the performance analysis of existing and hybrid algorithm in Fig. 5.

The performance obtained using Hybrid Privacy Preserving Scheme was found to be on top of the results obtained. The results depicts that Hybrid Privacy Preserving Scheme performs better than other considered classifiers models.

TABLE IV COMPARISON OFEXISTING AND HYBRID ALGORITHM

Parameters	K-means Clustering Algorithm	Decision Tree Algorithm	Hybrid privacy preserving scheme
Accuracy (%)	60.07	67.56	78.21
Sensitivity	0.427	0.407	0.306
Specificity	0.479	0.320	0.211
Fraudulent Rate (%)	43.93	43.34	29.52
Speed of detection	0.10	0.08	0.03

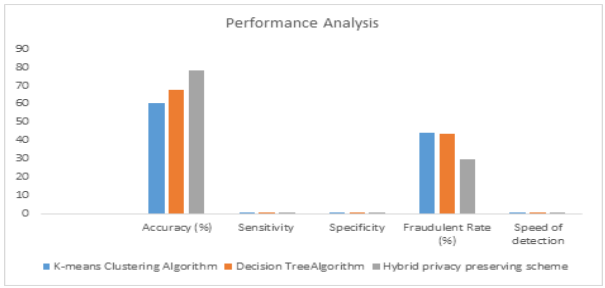

Fig. 5 Performance analysis of existing and hybrid algorithm

IV. CONCLUSION

Information mining procedures help clinician settle on appropriate choices in medicinal service applications. The Homomorphic algorithm uses probability based encryption to generate cipher text and Naive-bayes algorithm calculate prior probability which makes the encryption very efficient and fast. As a hybrid approach, these two algorithms are blended together and the results are analysed. Performance analysis are carried out using MATLAB tool. The hybrid

privacy preserving mechanism employing Naïve-Bayes and Homomorphic encryption algorithms classifies the data and provides security by keeping the patient data in the encrypted form by applying the respective algorithms.

Appendix A Dataset with 106 instances

Case #	Class	I0	PA500	HFS	DA	Area	A/DA	Max IP	DR	P
1	Car	524.79	0.19	0.03	228.80	6843.60	29.91	60.20	220.74	556.83
2	Car	330.00	0.23	0.27	121.15	3163.24	26.11	69.72	99.08	400.23
3	Car	551.88	0.23	0.06	264.80	11888.39	44.89	77.79	253.79	656.77
4	Car	380.00	0.24	0.29	137.64	5402.17	39.25	88.76	105.20	493.70
5	Car	362.83	0.20	0.24	124.91	3290.46	26.34	69.39	103.87	424.80
6	Car	389.87	0.15	0.10	118.63	2475.56	20.87	49.76	107.69	429.39
7	Car	290.46	0.14	0.05	74.64	1189.55	15.94	35.70	65.54	330.27
8	Car	275.68	0.15	0.19	91.53	1756.23	19.19	39.31	82.66	331.59
9	Car	470.00	0.21	0.23	184.59	8185.36	44.34	84.48	164.12	603.32
10	Car	423.00	0.22	0.26	172.37	6108.11	35.44	79.06	153.17	558.27
11	Car	410.00	0.32	0.30	255.82	10622.55	41.52	67.52	246.74	508.54
12	Car	500.00	0.23	0.05	219.30	9819.45	44.78	76.87	207.27	602.53
13	Car	438.78	0.21	0.06	120.90	4879.50	40.36	80.79	89.94	525.42
14	Car	366.94	0.28	0.25	172.75	7064.82	40.90	75.60	155.32	471.59
15	Car	485.67	0.23	0.13	253.89	8135.97	32.04	64.86	245.47	541.36
16	car	390.00	0.36	0.20	245.69	10055.84	40.93	70.32	236.49	477.55
17	car	269.50	0.21	0.04	80.41	1963.61	24.42	44.74	66.84	329.09
18	car	300.00	0.19	0.17	97.11	3039.56	31.30	51.35	82.42	387.08
19	car	325.00	0.22	0.29	229.22	5705.33	24.89	35.60	227.26	462.70
20	car	294.47	0.21	0.47	194.87	5541.26	28.44	36.77	191.80	445.51
21	car	500.00	0.19	0.19	144.69	3055.01	21.11	96.56	107.75	542.90
22	fad	211.00	0.05	0.09	30.75	151.98	4.94	14.27	27.24	217.13
23	fad	196.86	0.02	0.09	28.59	82.06	2.87	7.97	27.66	200.75
24	fad	245.00	0.19	0.08	62.90	1235.98	19.65	42.15	46.69	292.38
25	fad	352.66	0.12	0.09	68.53	1066.16	15.56	43.69	52.79	382.73
26	fad	243.29	0.04	0.07	68.54	383.93	5.60	9.99	67.82	263.64
27	fad	259.89	0.07	0.01	58.24	465.09	7.99	17.51	56.34	267.52
28	fad	250.00	0.07	-0.02	57.17	652.90	11.42	17.78	55.79	278.31
29	fad	200.00	0.04	0.12	42.32	220.81	5.22	10.68	40.95	218.03
30	fad	355.00	0.06	0.08	89.56	1033.85	11.54	27.56	86.58	372.04
31	fad	272.00	0.09	0.00	63.79	718.95	11.27	20.09	60.69	286.92
32	fad	341.62	0.09	0.07	85.04	1370.84	16.12	29.03	79.94	385.13
33	fad	160.32	0.18	0.16	37.22	341.88	9.19	30.89	20.76	187.57
34	fad	301.30	0.11	0.04	64.62	942.77	14.59	29.05	57.72	335.77
35	fad	155.00	0.17	0.12	38.94	415.11	10.66	25.84	29.13	184.82
36	fad	144.00	0.12	0.05	19.65	70.43	3.58	18.13	7.57	160.37
37	mas	178.00	0.17	0.21	41.54	489.44	11.78	35.75	21.16	215.91
38	mas				37.46	328.38	8.77		13.29	
39	mas	435.09	0.08	0.16	123.60	1342.28	10.86	37.38	117.81	433.20
40	mas	250.00	0.05	0.01	70.91	224.15	3.16	9.10	70.32	232.28
41	mas	339.51	0.05	0.03	88.63	331.08	3.74	19.83	87.62	307.79
42	mas	236.00	0.12	0.20	48.45	236.88	4.89	36.01	32.42	244.97
43	mas	481.47	0.08	0.02	79.06	1154.34	14.60	33.93	71.41	501.89
44	mas	252.00	0.11	0.03	38.54	493.79	12.81	25.54	28.87	280.66
45	mas	172.52	0.13	0.04	37.54	192.22	5.12	19.32	32.19	174.93
46	mas	121.00	0.17	0.09	24.44	144.47	5.91	22.02	10.59	141.77
47	mas	196.36	0.18	0.14	54.58	843.26	15.45	34.15	42.58	239.94

Case #	Class	I0	PA500	HFS	DA	Area	A/DA	Max IP	DR	P
48	mas	370.40	0.10	0.00	115.92	1308.12	11.28	31.37	112.72	365.98
49	mas	260.28	0.08	0.03	58.82	277.26	4.71	17.87	56.04	248.62
50	mas	544.65	0.06	0.00	100.79	1189.29	11.80	29.41	96.58	553.36
51	mas	310.00	0.17	0.17	98.51	2741.03	27.82	49.33	85.27	388.98
52	mas	274.99	0.15	0.14	66.46	1217.42	18.32	40.85	52.42	327.56
53	mas	281.32	0.23	0.44	157.88	5305.12	33.60	46.38	150.92	398.90
54	mas	327.00	0.14	0.08	76.21	1664.67	21.84	43.22	62.77	379.26
55	gla	470.52	0.13	0.07	150.22	2657.91	17.69	47.56	142.50	491.47
56	gla	223.00	0.12	0.08	33.10	197.01	5.95	30.45	12.96	252.48
57	gla	152.00	0.17	0.23	34.22	94.35	2.76	31.28	13.88	180.61
58	gla	303.00	0.06	0.04	22.57	102.50	4.54	21.83	5.72	321.65
59	gla	250.00	0.09	0.09	29.64	180.76	6.10	26.14	13.96	280.12
60	gla	197.00	0.13	0.07	33.46	409.65	12.24	26.99	19.77	231.78
61	gla	197.00	0.13	0.07	33.46	409.65	12.24	26.99	19.77	231.78
62	gla	216.41	0.12	0.07	53.60	280.45	5.23	22.79	48.51	215.37
63	gla	178.00	0.15	0.10	40.29	474.40	11.77	25.92	30.85	209.18
64	gla	185.00	0.15	0.09	39.89	361.75	9.07	26.86	29.49	210.18
65	gla	391.00	0.06	0.01	35.78	265.15	7.41	22.13	28.11	400.99
66	gla	502.00	0.07	0.03	53.24	834.27	15.67	33.33	41.51	544.04
67	gla	176.00	0.09	0.08	20.59	79.71	3.87	18.23	9.58	191.99
68	gla	145.00	0.12	0.11	21.22	82.46	3.89	20.30	6.17	162.51
69	gla	124.13	0.13	0.11	20.59	78.34	3.80	18.46	9.12	134.89
70	gla	103.00	0.16	0.29	23.75	78.26	3.29	22.32	8.12	124.98
71	con	1724.09	0.05	-0.02	404.13	3053.97	7.56	71.43	399.19	1489.39
72	con	1385.66	0.09	0.09	202.48	8785.03	43.39	143.09	143.26	1524.61
73	con	1084.25	0.07	0.00	191.90	2937.97	15.31	66.56	179.98	1064.10
74	con	649.37	0.11	0.02	207.11	3344.43	16.15	50.55	200.85	623.91
75	con	1500.00	0.06	0.05	375.10	4759.45	12.69	78.45	366.80	1336.16
76	con	770.00	0.04	0.00	175.02	346.09	1.98	25.22	173.19	654.80
77	con	650.00	0.04	0.15	216.81	427.53	1.97	33.77	214.17	528.70
78	con	691.97	0.03	0.09	190.68	304.27	1.60	23.98	189.16	594.32
79	con	1461.75	0.04	0.05	391.85	5574.00	14.22	57.23	387.64	1428.84
80	con	1496.74	0.10	0.08	640.28	11072.00	17.29	108.29	631.05	1178.27
81	con	1111.81	0.10	0.07	386.99	7659.74	19.79	86.03	377.30	990.98
82	con	1270.67	0.08	0.07	555.35	3612.97	6.51	68.78	551.08	895.19
83	con	1647.94	0.08	0.09	576.77	11852.49	20.55	111.44	565.90	1402.88
84	con	1535.85	0.09	0.00	637.35	10814.05	16.97	96.61	632.17	1197.76
85	adi	2100.00	0.06	-0.05	390.48	16640.72	42.62	125.90	380.64	2073.03
86	adi	1800.00	0.03	0.04	301.06	4406.15	14.64	67.63	293.37	1742.38
87	adi	2100.00	0.12	0.38	450.55	35671.61	79.17	436.10	113.20	2461.45
88	adi	1666.15	0.01	0.06	72.93	1402.23	19.23	51.85	58.60	1746.58
89	adi	1700.00	0.04	0.11	120.65	12331.10	102.20	120.30	-9.26	2212.18
90	adi	1949.12	0.05	0.02	170.33	3212.08	18.86	101.46	136.82	1941.37
91	adi	1850.00	0.08	0.07	253.62	13113.20	51.70	160.07	196.73	1916.99
92	adi	2350.00	0.08	0.27	515.29	27758.64	53.87	289.57	426.23	2457.68
93	adi	1800.00	0.09	0.21	362.86	15021.55	41.40	217.83	290.20	1893.66
94	adi	1900.00	0.05	0.11	272.62	7481.59	27.44	138.36	234.90	1924.52
95	adi	1800.00	0.07	0.16	385.56	13831.72	35.87	157.57	351.90	1823.03
96	adi	1850.00	0.07	0.23	325.19	8644.98	26.58	208.74	249.35	1908.18
97	adi	1650.00	0.05	0.04	274.43	5824.90	21.23	81.24	262.13	1603.07
98	adi	2800.00	0.08	0.18	583.26	31388.65	53.82	298.58	501.04	2896.58

Case #	Class	I0	PA500	HFS	DA	Area	A/DA	Max IP	DR	P
99	adi	2329.84	0.07	0.35	377.25	25369.04	67.25	336.08	171.39	2686.44
100	adi	2400.00	0.08	0.22	596.04	37939.26	63.65	261.35	535.69	2447.77
101	adi	2000.00	0.07	0.12	330.27	15381.10	46.57	169.20	283.64	2063.07
102	adi	2000.00	0.11	0.11	520.22	40087.92	77.06	204.09	478.52	2088.65
103	adi	2600.00	0.20	0.21	1063.44	174480.5	164.07	418.69	977.55	2664.58
104	adi	1600.00	0.07	-0.07	436.94	12655.34	28.96	103.73	432.13	1475.37
105	adi	2300.00	0.05	0.14	185.45	5086.29	27.43	178.69	49.59	2480.59
106	adi	2600.00	0.07	0.05	745.47	39845.77	53.45	154.12	729.37	2545.42

REFERENCES

[1] Mehrnoosh, Monshizadeh and Zheng Yan, "Security Related Data Mining", *IEEE International Conference on Computer and Information Technology*, pp. 775 – 782, 2014.

[2] Dileep Kumar Singh and Vishnu Swaroop, "Data Security and Privacy in Data Mining: Research Issues & Preparation", *International Journal of Computer Trends and Technology*, Vol. 4, No.2, pp. 194-200, 2013.

[3] E Bertino, I.N. Fovino and L.P Provenza, , "A Framework for Evaluating Privacy Preserving Data Mining Algorithms", *Data Min Knowledge Disc*, Vol. 11, No. 2, pp. 121–154, 2005.

[4] Kemal polat and Salih Gunes, "A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems", *Expert Systems with Applications, Elsevier*, Vol. 36, No. 2, Part 1, pp. 1587-1592, 2009.

[5] R.W.K Leung, H.C.W. Lau and C.K. Kwong, "On a responsive replenishment system: a fuzzy logic approach", *Expert Systems*, Vol. 20, pp. 20-32, 2003.

[6] A Inokuchi, T Washio and H Motoda, "An apriori-based algorithm for mining frequent substructures from graph data," *Proceedings of the Fourth European Symposium on the Principle of Data Mining and Knowledge Discovery, Lyon, France*, pp.13-23, 2000.

[7] Mohammad Khubeb, Siddiqui and Shams Naahid, *International Journal of Database Theory and Application*, Vol. 6, No. 5, pp. 23-34, 2013.

[8] CC Aggarwal and PS Yu, "A General Survey of Privacy-Preserving Data Mining Models and Algorithms", *Springer, Boston,MA* Vol. 34, pp. 11-52, 2008.

[9] J.B. Awotunde, O.E. Matiluko and O.W Fatai , "Medical Diagnosis System Using Fuzzy Logic", *Afr J Comp and ICT*, Vol. 7. No.2, pp. 99-106, 2014.

[10] S Das, D Guha, and B Dutta, "Medical diagnosis with the aid of using fuzzy logic and intuitionistic fuzzy logic", *Appl. Intell.*, Vol. 45, No. 3, pp. 850–867, 2016.

[11] K.M. Al-Aidaroos, A.A. Bakar and Z. Othman, "Medical Data Classification with Naive Bayes Approach", *Information Technology Journal*, Vol. 11, pp.1166-1174, 2012.

[12] Tung-Shou Chen, Jeanne Chen and Yuan-Hung Kao, "A Novel Hybrid Protection Technique of Privacy-Preserving Data Mining and Anti-Data Mining", *Information Technology Journal*, Vol. 9, pp. 500-505, 2010.

[13] V. Subramaniyaswamy and S. Chenthur Pandian, "A Complete Survey of Duplicate Record Detection Using Data Mining Techniques", *Information Technology Journal*, Vol. 11 pp. 941-945, 2012.

[14] S Dhanashree Medhekar, Mayur P Bote, Shruti D and Deshmukh, "Heart Disease Prediction System using Naive Bayes", *International Journal of Enhanced Research in Science Technology and Engineering*, Vol. 2, No. 3, pp. 1-5, 2013.

[15] J Jossinet and Physiol Meas, "The impedivity of freshly excised human breast tissue", Vol. 19, No. 1, pp. 61-75, 1998.

[16] K Julie, MD Taitsman, Christi Macrina, M.P.A Grimm and MD Shantanu Agrawal, "Protecting Patient Privacy and Data Security", *The New England Journal Of Medicine*, Vol. 14, pp. 977-979, 2013.

[17] Adriana Lopez-Alt, Eran Tromer and Vinod Vaikuntanathan, "On-the-Fly Multiparty Computation on the Cloud via Multikey Fully Homomorphic Encryption", ISBN: 978-1-4503-1245-5NY, USA, : Proceeding STOC '12 Proceedings of the forty-fourth annual ACM symposium on Theory of computing Pp. 1219-1234, 2013.

[18] Mohammad Khubeb Siddiqui and Shams Naahid, "Analysis of KDD CUP 99 Dataset using Clustering based Data Mining", *International Journal of Database Theory and Application*, Vol. 6, No. 5, pp. 23-34, 2013.

[19] Y. Dhanalakshmi and Dr I, Ramesh Babu, "Intrusion Detection Using Data Mining Along Fuzzy Logic and Genetic Algorithms", *IJCSNS International Journal of Computer Science and Network Security*, Vol. .8, No. 2, pp. 27-32, 2008.

[20] Rahul Deo Sah and Dr Jitendra Sheetalani, "Review of Medical Disease Symptoms Prediction Using Data Mining Technique", *IOSR Journal of Computer Engineering (IOSR-JCE)*, Vol. 19, No. 3, Ver. I, pp. 59-70, 2017.

[21] J Kamber, M. Han and J Pei, "Data Mining Concepts and Techniques. Morgan Kaufmann Publishers", San Francisco. 2012.

[22] Mevlut Ture, "Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients", *Elsevier*, Vol. 36, No. 2, pp. 2017-2026, 2009.

[23] Pradeep Kumar, Kishore Indukuri Varma and Ashish Sureka,"Fuzzy based clustering algorithm for privacy preserving data mining", Int. J. Business Information Systems, Vol. 7, No. 1, pp 27-40, 2011.

[24] A.R. Aronson, "Effective mapping of biomedical text to the UMLS Meta thesaurus: the Meta Map program", *Proc. AMIA Symp.*, pp.17–21, 2001.

[25] A. Sung and S. Mukkamala, "Identifying important features for intrusion detection using SVM and neural networks", *in symposium on application and the Internet*, pp. 209-216, 2003.