International Journal of Computer Applications Technology and Research
Volume 5– Issue 8, 509 - 515, 2016, ISSN:- 2319–8656

A Gene Structure Prediction Model using Bayesian Algorithm and the Nearest Neighbor

Elham Naseh
Department of Computer Engineering, Qeshm international Branch, Islamic Azad University
Qeshm ,Iran

*Ali Asghar Safaee
Department of Medical Informatics,
Faculty of Medicine,
Tarbiat Modarres University
Tehran, Iran

Abstract: Basically genetic disorders include general problems and issues that are caused by the failure of one or more of the genome, and usually appear at birth; although they sometimes occur later. Genetic diseases may not be inherited and they may be caused as new mutations in the genome of embryos. Like many other diseases, diagnosis, treatment, and prognosis of genetic diseases is very important and sometimes complex. One of the best ways of treating genetic diseases is its diagnosis in a fetus. All gene structures of a fetus should be available in order to diagnose genetic disease. This structure can be achieved when the fetus is seven months and there are just 5% to 30% of gene sequence structure before seven months. To solve this shortcoming and fix the obstacle in the diagnosis of diseases, 5 to 30% of the gene sequence structure of the whole structure of the fetus is predicted with the help of parents' gene structure. In previous studies, gene structure prediction using machine learning algorithms has achieved a maximum accuracy of 95%.

Keywords: Gene structure prediction, nearest neighbor algorithm, Bayesian algorithm, blended learning methods, genetic disease diagnosis.

1. INTRODUCTION

Genetic diseases are one of the most serious issues in the field of health and medicine and researchers' efforts in related sciences and various fields still continues. Genetic diseases include diseases that are caused by failure or mutations in the genes or genetic material of humans. These diseases often occur at birth, but can also occur years later. Genetic diseases may not be inherited, for example they may be created as a result of new mutations in the genome of the fetus.

If genetic diseases are diagnosed before birth, they are treatable. One of the ways to diagnose genetic disease is based on the availability of fetus' genetic sequence. But the challenge is that only about 10% of genes have been identified in the embryonic time and 90% remains uncertain. In other words, the entire structure of genes is completed at birth, when almost all genetic diseases are no longer curable. Therefore, the remaining uncertain 90% of fetus' genetic structure should be predicted with the help of parents' chromosome structure and 10% of fetus' cell sequence in order to overcome this challenge [8, 9]. As it seems, predicting something that much of it is not available based on a much smaller proportion (predicting remaining 90% based on 10%) is very difficult and acceptable accuracy in the prediction will be of great importance.

To learn more about the issue that we face, some additional details are provided here. Deoxyribonucleic Acid (DNA), discovered in 1869 by Friedrich Miescher [1], is a chemical structure that creates chromosomes. The part of chromosome that has unique characteristics is called a gene.

DNA has a spiral structure and is formed by a double-stranded genetic material, wrapped around each other in a spiral form. Each strand contains a base composition that is called nucleotide. Each basic composition is formed by four structures (adenine A, guanine G, cytosine C, thymine T). Human cells have twenty-three pairs of chromosomes [2, 3].

Human cells inherit two different groups of chromosomes, one group from father and the other from mother. Each group of chromosome is made up of 23 single chromosome (22 asexual chromosomes and one sexual chromosome, either in the form of Y or in the form of X). The group of chromosome that can be seen in Figure 1 is for father's chromosomes (since it has chromosome X (XY) and chromosome Y). If the group of chromosome is for mother, the group will include chromosome X and again chromosome X.

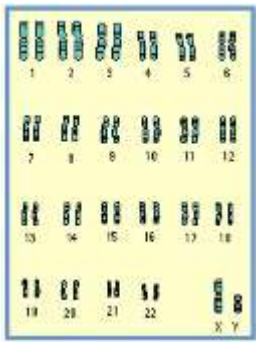

Figure 1: Father's chromosome structure [4]

In principle, every living entity has many attributes of: color, size, horned or polled. These are a few of the traits that are passed from parents to offspring. These traits are controlled by the genes. Genes are small and complex molecules that are on chromosomes. Any existing entity contains several thousand pairs of genes, half of which is inherited from father and the other half is inherited from mother, and these genes together form genotype of that existing entity [5].

Genotype determines the approximate function of an organism on different characteristics. Each pair of genes control some of the characteristics. Some characteristics have a very high heritability, but some others have a very low heritability [6, 7].

In this case, the gene structure of a fetus should be predicted with the help of parents' genes. Gene structure is a series of

DNA coding sequence. DNA has a basis code of two. DNA structures can be seen in Table 1.

Table 1: DNA code structure

Gene name	Equivalent code	Equivalent binary code
Guanine (G)	0	00
Cytosine (C)	1	01
Adenine (A)	2	10
Thymine (T)	3	11

Human beings have a sequence of DNA. An example of sequence can be seen in Figure 2.

DNA code	C	A	C	C	T	T	G	G	C	T	T	C	C
Equivalent code	1	2	1	1	3	3	0	0	1	3	3	1	1
Equivalent binary code	01	10	01	01	11	11	00	00	01	11	11	01	01

Figure 2: Gene sequence structure

Gene sequence of a fetus for any DNA is one of the following three conditions:

• Father's code has been copied.

• Mother's code has been copied.

• None of mother's or father's code; a mutation has occurred in this case.

In Figure 3, the gene structure of a fetus, father, and mother has been shown using DNA code.

	1	2	3	4	5	6	7	8	9	10	11	12	13
Father's DNA code	C	A	C	C	T	T	G	G	C	T	T	C	C
Mother's DNA code	A	A	C	T	T	A	T	G	A	C	G	C	T
Fetus' DNA code	A	A	C	C	T	T	G	G	A	A	T	C	C

Figure 3: Fetus' DNA structure based on parents' DNA

As can be seen in Figure 3, each DNA contains 13 gene sequence. Each gene sequences that include father, mother, and fetus' code is called the Trio. In Figure 3, 13 Trio can be seen, In Trios one to three, father's DNA code is copied, mother's DNA code is copied in Trios eight to nine, gene mutation has occurred in Trio ten, and father's DNA is copied in Trios eleven to thirteen.

The change of DNA code from father to mother and vice versa is called crossover. Crossover has occurred in trios three to four, seven to eight, nine to ten, and ten to eleven.

Up to 5 to 30% of a fetus' DNA codes are determined up to 7 month and the remaining DNA codes are changing. An example of codes excluded from DNA structure can be seen in Figure 4.

	1	2	3	4	5	6	7	8	9	10	11	12	13
Father's DNA code	01	10	01	01	11	11	00	00	01	11	11	01	01
Mother's DNA code	10	10	01	11	11	10	11	00	10	01	00	01	11
Fetus' DNA code	10	?	?	?	?	11	?	?	?	?	?	?	01

Figure 4: Equivalent binary code of sequence structure of fetus and father and mother in figure 3 that just 3 trios are determined

As can be seen in Figure 4, only 3 Trios of the fetus have been found, but fetus' parents' codes are all determined. Now all fetus' codes should be identified with the help of prediction models (such as the proposed algorithm).

The rest of the article would be as follows: the research papers and related work are introduced in section 2. The proposed model to predict gene structure is presented in section 3 and section 4 empirically evaluates the proposed model. Conclusions and recommendations for future studies are provided in sections 5 and 6.

2. RESEARCH BACKGROUND

In a study conducted by Rutkoski et al. (2013) entitled as "imputation of unordered markers and the impact on genomic selection accuracy", it was shown that genomic selection is a breeding method to accelerate the rate gene gain. In this study, using four empirical datasets, four imputation methods including k-nearest neighbors, singular value decomposition, random forest regression, and expectation maximization imputation were evaluated and characterized in terms of their imputation accuracies and the factors affecting accuracy. It was shown that SMV has a high accuracy of 93%.

In 91 years, Mrs. Maryam Ali in his thesis on DNA and predicted protein structure, and neuro-genetic algorithm used to predict [10], the results show that these methods increase the overall accuracy, quality improvement predict and solve the problem of unbalanced data, the prediction accuracy of 92% have been reported in this study.

Alireza (2012) predicted protein structure on DNA using Neuro-genetic algorithm in her thesis [10]. The results show that these methods increase the overall accuracy, improve prediction quality, and solve the problem of unbalanced data. The prediction accuracy has been reported to be 92% in this study.

In this study, we set out to achieve higher accuracy in the proposed algorithm; therefore, a combination of Bayesian algorithm and nearest neighbor has been used.

Bayesian algorithm is a statistical method that is based on the conditional probability, i.e. by classifying we seek to define the class of a new sample, which is the main goal of classification [17].

In this method, which is based on probability and statistical methods, it is found that how probable a sample belongs to a specific class and it is finally concluded that which class does the sample belong based on the highest obtained probability. Finding out the probability is the main challenge of Bayesian-based methods [18].

K-nearest neighbor learning algorithm is one of the most famous algorithms in the field of learning, the performance and features of which will be discussed [10].

In sample-based methods or sample-based learnings, classification of a new sample is done in a way that the new sample is compared with all the samples of a class and then the samples that are more similar to the new sample are extracted as the ones with the potential to have the same class with the new sample and then the class of the new sample is determined on the basis of some methods.

In the nearest neighbor, K of the nearest neighbors are determined from the training set. Now if K is equal to one, one of the nearest neighbors is extracted for the new sample and its class is investigated. The value of the nearest neighbor will be the same as the class of the new sample.

In [10], 90% of gene structure is predicted using a combination of two algorithms of support vector machine and neural networks, and structure of the parents' chromosomes as well as the 10% of the fetus' cell sequence. Support vector

machine algorithm has a good performance in dealing with linear problems and artificial neural network is useful for nonlinear problems. Thus, since predicting 90% of the structure is both linear and non-linear, the combination of two algorithms was used [10].

The proposed model to predict gene structure is presented in the next chapter and them its performance will be evaluated.

3. THE PROPOSED MODEL TO PREDICT GENE STRUCTURE

In order to provide the proposed prediction model, the problem should be first categorized using the nearest neighbor and Bayesian approach, then a new method is provided to combine two nearest neighbor method and the Bayesian approach to solve the problem, which will be explained in part 3 -3 in detail.

In predicting fetus' gene structure, the data, in principle, contains parents' and fetus' features that should be predicted with the help of parents' information, fetus' status, and his genetic structure. In order to achieve a category with a better accuracy with both methods, features of the table number are added to the information. Features of table number starts from 1 and has a rising rate. The reason for adding features of the table number is to determine the difference between each row. For example, first row is closer to the second row than the tenth row, because the difference between first row and tenth row is ten units, but the difference between the first row and the second row is two units. The information by adding the numbers of the table can be seen in Figure 5.

No.	Father	Mother	(fetus) Class
1	01	10	**10**
2	10	10	**10**
3	01	01	**01**
4	01	11	**01**
5	11	11	**11**
6	11	10	**11**
7	00	11	**00**
8	00	00	**00**
9	01	10	**10**
10	11	01	**10**
11	11	00	**11**
12	01	01	**01**
13	01	11	**01**

Figure 5: Adding features of number to each Trio

3.1 Prediction model based on the nearest neighbor

Since in this method values are binary, there is no need for pre-processing such as dimension reduction or numbers' transformation. To implement the nearest neighbor algorithm, the following steps are implemented:

1. First, the data are divided into two sets.

   1.1. The data set that its class label is determined (the training data set).

   1.2. The data set that its class label is not determined (the test data set).

2. The following steps are taken for each of the unlabeled data.

   2.2 Euclidean distance of each data from all labeled data is calculated.

   2.2. K nearest distance is selected as the most similar.

3. The class label is determined from the most similar Trios using majority voting rule.

4. The accuracy of the labels specified in Step 2 is calculated.

5. Algorithm runs from 1 to 10 for K, and the best value of K in terms of accuracy is determined as output.

Dividing data into two training and test sets is shown in Figure 6. Then the Euclidean distance of each test Trio from training Trio is calculated. Given that the labeled class data is discrete, majority voting method is used to select K-nearest training data and then test data class label is determined using voting method.

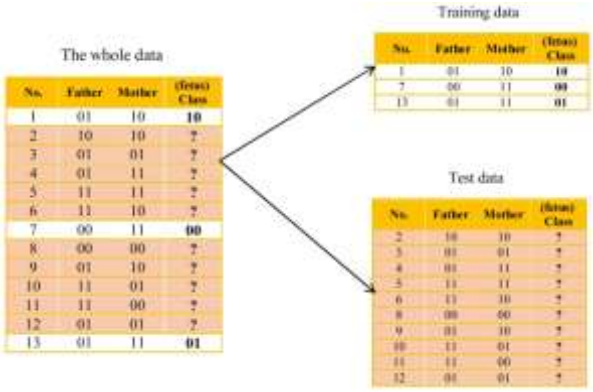

Figure 6: Dividing data into two training and test sets

3.2 Prediction Model Based on Bayesian Method

The issue of detecting fetus has four different classes. The following steps are taken to run the algorithm:

1. First, data is divided into two sets.

   1.1. The data set that its class label is determined.

   1.2. The data set that its class label is not determined.

2. The following steps are taken for each of the unlabeled data.

   2.1. The extent to which data belongs to each one of the four classes is calculated.

   2.2 The maximum amount is selected as the label.

3. The accuracy of the labels specified in Step 2 is calculated.

The first step in Bayesian algorithm is shown in Figure 6. The following equation is used to calculate the value of four classes.

$$\begin{cases} P(00|X) = P(x_1|00) \times P(x_2|00) \times P(x_3|00) \times P(00) \\ P(01|X) = P(x_1|01) \times P(x_2|01) \times P(x_3|01) \times P(01) \\ P(10|X) = P(x_1|10) \times P(x_2|10) \times P(x_3|10) \times P(10) \\ P(11|X) = P(x_1|11) \times P(x_2|11) \times P(x_3|11) \times P(11) \end{cases}$$

Then the probability of each above relationships is calculate and the highest probability is selected as the label.

3.3 Hybrid model

In order to combine the two Bayesian and the nearest neighbor methods to achieve a higher accuracy than either methods, a hybrid model is presented as follows.

The proposed algorithm is in a way that data is first divided into two training and test sets. Then, the nearest neighbor and Bayesian algorithms are implemented. The outputs of the test data are investigated and if both algorithms have the same outputs, the set is added to the training data, but if the two algorithm does not have the same output, the set will remain in the test data. Then both algorithms are run with new data of training set and the previous trend is done for test data. This procedure is repeated so that one of the two following states happens.

• Test data will end.

• No element of the training set is added to the test set.

If the first case occurs, then the algorithm has completed and the initial test set is evaluated. But if the latter occurs, the nearest neighbor method will be chosen as the output. The nearest neighbor method is chosen because the accuracy of the nearest neighbor method is higher than the Bayesian approach.

The proposed algorithm is shown in Figure 7.

```
Function Accuracy=Classification(input =Gen)
    1.  Cross Validation Gen To Training & Test
    2.  While (True)
    3.      X ← Classify Bayesian (Training)
    4.      Y ← Classify KNN (Training)
    5.      A ← Evaluation X (Test)
    6.      B ← Evaluation Y (Test)
    7.      If (The number of outputs A and B are the same)
    8.          The same output will be removed from the test set and the
                training set is added.
    9.      Else
    10.         Exit While
    11.     End IF
    12. End While
    13. If No output is not same
    14.     Classify KNN (Training)
    15. End if
    16.     Accuracy ← Evaluation (Test)
End
```

Figure 7: Pseudo code of the proposed algorithm

As can be seen in figure 7, data is divided into two training and test sets in the first line of the proposed pseudo code without any pre-processing, because the data is binary and there is no need for transformation. That is, data is complete and there is no missing or extreme data. It should be noted that data has four features of No, father's code, mother's code, and fetus' code (Class Label) and there is no need for dimension reduction.

Lines 2 to 12 are the main loops of this program. Two new models are developed using the nearest neighbor and Bayesian approach in lines 3 and 4 of the training data and then these two models are evaluated in lines 5 and 6 using test data. Finally, in line 7 it will be checked whether the answer to the two models is the same or not?

If a part of the test elements is similar, then line 8 is run. At this stage, a number of test data elements that have the same answer are removed from the test data sets and added to the training data set. However, if all the elements are the same or none of the elements are the same, the program exits the main loop on line 10.

If the output of all the test data is not the same in both the nearest neighbor and the Bayesian approaches in lines 13 to 15, the nearest neighbor is taken into account and then the accuracy of test data is sent as output in line 16.

4. EXPERIMENTAL EVALUATION

One of the most important parts of a theory is doing the experiments and proving their results. In order to test the proposed theory, some programs were created using MATLAB that will be explained and displayed here. The experiments were conducted on a computer with a 4GH processor and 6GB ram. The real data has been used to test data, which was collected from medical sciences of Mashhad and included 2 million records from parents' and fetus' gene. The data was stored on a 2 basis and a view of the data is shown in Figure 8.

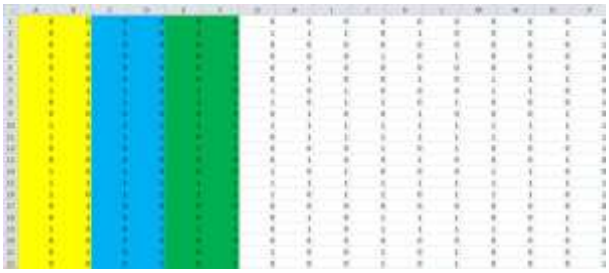

Figure 8: gene sequence data to evaluate the proposed algorithm

As shown in Figure 8, the first two columns are father's code, the second two columns are mother's code, and the last two columns are fetus' code. The missing data of fetus' gene is a number between 70 to 95%; therefore, 70 to 95% of the data must be removed.

The four parameters of Confusion Matrix, Sensitivity, Specificity and Accuracy are examined, then the time for their implementation will also be examined. The output can be seen in Table 1.

Confusion Matrix shows the performance of the algorithms. Usually this performance is used for algorithms such as decision tree. Each column of the matrix shows an example of the predicted value. If each row has an actual (right) sample, the structure of Confusion Matrix is seen in Figure 9.

		Predicted Class: Yes	Predicted Class: No
Actual Class	Yes	TP	FN
	No	FP	TN

Figure 9: Confusion Matrix structure

The formulas for Sensitivity, Specificity, and Accuracy are calculated with the help of Confusion Matrix that can be seen in equations (1), (2), and (3).

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (2)$$

$$\text{Accuracy} = \frac{TN+TP}{FP+TN+TP+FN} \quad (3)$$

Missing value of data ranges between 70 and 95% [1]. If we call the missing value K, K% of data should first be deleted randomly. Then, both the nearest neighbor algorithm and

Bayesian algorithm are run once, and then confusion matrix is calculated for both algorithms. With the help of confusion matrix, the three parameters of Sensitivity, Specificity and Accuracy are calculated. With the exception of the confusion matrix, execution time of both algorithms is calculated. As can be seen in Table 2, the nearest neighbor algorithm has a higher accuracy than Bayesian algorithm, but the nearest neighbor algorithm's execution time is three times more than Bayesian algorithm.

4.1 Comparison of the proposed hybrid algorithm and the nearest neighbor approach

Given that the nearest neighbor algorithm has better accuracy than the Bayesian approach; therefore, the proposed algorithm has been evaluated with the nearest neighbor method that can be seen in Table 3 and Figure 12. As can be seen in Table 3, the measure of the accuracy of the proposed algorithm has been better than the nearest neighbor algorithm. In all the missing values, the accuracy has increased; the minimum accuracy increase was in 73% of missing values in a way that accuracy had increased .17% and the maximum accuracy increase was in 95% of missing values in a way that accuracy had increased 1.68%. On average, an increase of .7% has happened.

Table 3: Comparison of the nearest neighbor algorithm and the proposed algorithm based on accuracy

ID	Missing value	Nearest Neighbor	proposed algorithm
		Accuracy	Accuracy
1	70	98.00%	98.54%
2	71	98.03%	98.32%
3	72	97.78%	98.12%
4	73	97.81%	97.89%
5	74	97.57%	97.93%
6	75	97.60%	97.91%
7	76	97.50%	97.90%
8	77	97.41%	97.81%
9	78	97.19%	97.80%
10	79	97.22%	97.73%
11	80	97.00%	97.62%
12	81	97.04%	97.53%
13	82	96.83%	97.34%
14	83	96.63%	97.32%
15	84	96.42%	97.31%
16	85	96.24%	97.01%
17	86	96.04%	96.86%
18	87	95.75%	96.74%
19	88	95.46%	96.49%
20	89	95.17%	96.12%
21	90	94.78%	95.32%
22	91	94.30%	95.31%
23	92	93.82%	95.07%
24	93	93.33%	94.24%
25	94	92.55%	93.72%
26	95	91.85%	93.53%

5. CONCLUSION AND SUGGESTIONS FOR FURTHER STUDIES

Gene structure prediction is one of the most important ways to diagnose genetic diseases. These diseases often occur at birth, but can also occur years later. Genetic diseases may not be inherited, for example they may be created as a result of new mutations in the genome of the fetus. In this study, it was shown that 95% of missed genes is predictable. It was also found that a model composed of the nearest neighbor algorithm and Bayesian algorithm is used to predict missing data of genes.

The proposed model is a combination of the nearest neighbor algorithm and Bayesian algorithm. The structure of this algorithm is based on voting and sameness of the two algorithms' output and it also has a higher accuracy than the nearest neighbor algorithm and the Bayesian algorithm.

In the proposed nearest neighbor algorithm, it was showed that the algorithm has high capability in solving this problem, and it was also revealed that the implementation of this algorithm is very time consuming, but has a high accuracy.

Bioinformatics algorithms must be appropriate in terms of Sensitivity and Specificity. If the value of one of these two parameters is low, the algorithm is not acceptable. In the present study, it was shown that a combination of several algorithms can be used to optimize both parameters.

Past methods have an accuracy of about 90% through which lots of diseases were predictable, but the proposed algorithm has reached an accuracy of 98%, which is more accurate than other methods. Therefore, it can give a better prediction in the structure of the gene sequence.

To continue the work done in identifying gene sequences, some suggestions are put forward:

• Parents' gene sequence was only used in the proposed system, but if the gene sequence of grandparents or siblings is used, better results can be achieved.

• Non-linear algorithms such as decision tree can be used instead of Bayesian algorithm.

• Reinforcement learning algorithms can be used to adjust the algorithm parameters (such as the K value in the nearest neighbor).

• The use of neuro-fuzzy networks can probably bring about interesting results in diagnosis.

6. REFERENCES

[1] Langman's Medical Embryology, Thirteenth, North American Edition edition. Philadelphia: LWW, 2014.

[2] Larsen's Human Embryology, 5e, 5 edition. Philadelphia, PA: Churchill Livingstone, 2014.

[3] BRS Embryology, Sixth edition. LWW, 2014.

[4] High-Yield Embryology, Fifth edition. Philadelphia: LWW, 2013.

[5] The Developing Human: Clinically Oriented Embryology, 10e, 10 edition. Philadelphia, PA: Saunders, 2015.

[6] Developmental Biology, Tenth Edition, 10 edition. Sunderland, MA, USA: Sinauer Associates, Inc., 2013.

[7] J. E. Rutkoski, J. Poland, J.-L. Jannink, and M. E. Sorrells, "Imputation of unordered markers and the impact on genomic selection accuracy," G3 Genes Genomes Genet., vol. 3, no. 3, pp. 427–439, 2013.

[8] H. Pang, M. Hauser, and S. Minvielle, "Pathway-based identification of SNPs predictive of survival," Eur. J. Hum. Genet., vol. 19, no. 6, pp. 704–709, 2011.

[9] J. He, J. Zhang, G. Altun, A. Zelikovsky, and Y. Zhang, "Haplotype tagging using support vector machines.," in GrC, 2006, pp. 758–761.

[10] D. Gutierrez, Machine Learning and Data Science: An Introduction to Statistical Learning Methods with R, First edition. Technics Publications, 2015.

[11] Machine Learning in Action, 1 edition. Shelter Island, N.Y: Manning Publications, 2012.

[12] M. ACI, DEVELOPMENT OF TWO HYBRID CLASSIFICATION METHODS FOR MACHINE LEARNING: Using Bayesian, K Nearest Neighbor Methods and Genetic Algorithm. LAP LAMBERT Academic Publishing, 2011.

[13] J. Kleissl, Solar Energy Forecasting and Resource Assessment, 1 edition. Academic Press, 2013.

[14] Doing Data Science: Straight Talk from the Frontline, 1 edition. Beijing ; Sebastopol: O'Reilly Media, 2013.

[15] "Data Mining in R - Learning with Case Studies" taught by Dr. Luis Torgo - 4 week online course - 6/28/13 to 7/26/13. .

[16] Machine Learning, 1 edition. New York: McGraw-Hill Education, 1997.

[17] P. Bessiere, E. Mazer, J. M. Ahuactzin, and K. Mekhnacha, Bayesian Programming, 1 edition. CRC Press, 2013.

[18] M.-H. Chen, L. Kuo, and P. O. Lewis, Eds., Bayesian Phylogenetics: Methods, Algorithms, and Applications. CRC Press, 2014.

[19] P. Domingos, The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books, 2015.

[20] Pattern Recognition and Machine Learning. New York: Springer, 2007.

[21] A First Course in Bayesian Statistical Methods, 1st ed. 2009 edition. London ; New York: Springer, 2009.

[22] Bayesian Analysis for the Social Sciences, 1 edition. Chichester, U.K: Wiley, 2009.

[23] Machine Learning for Healthcare, 1 edition. O'Reilly Media, 2015.

[24] Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference, 1 edition. New York: Addison-Wesley Professional, 2015.

[25] Bayesian Estimation and Tracking: A Practical Guide, 1 edition. Hoboken, N.J: Wiley, 2012.

[26] Bayesian Reasoning and Machine Learning, 1 edition. Cambridge ; New York: Cambridge University Press, 2012.

Table 2: Comparison of the two the nearest neighbor and Bayesian algorithms

		The nearest neighbor algorithm						Bayesian algorithm					
Number	**Missing value**	Confusion Matrix		Sensitivity	Specificity	Accuracy	Time/sec.	Confusion Matrix		Sensitivity	Specificity	Accuracy	Time/sec.
1	70	434	7	98.00%	97.99%	98.00%	55.762006	302	49	97.78%	98.02%	97.90%	2.083503
		7	342					48	300				
2	71	348	7	98.03%	98.02%	98.03%	55.251918	307	49	97.78%	98.02%	97.90%	2.022653
		7	347					49	305				
3	72	352	8	97.78%	97.78%	97.78%	56.160841	311	51	98.09%	97.74%	97.92%	1.895628
		8	352					49	309				
4	73	357	8	97.81%	97.80%	97.81%	57.503536	315	51	97.57%	97.49%	97.53%	1.862098
		8	356					50	313				
5	74	362	9	97.57%	97.57%	97.57%	59.087744	320	52	97.59%	97.34%	97.46%	1.749257
		9	361					51	318				
6	75	366	9	97.60%	97.60%	97.60%	55.781443	323	52	97.59%	97.34%	97.46%	1.911674
		9	366					62	322				
7	76	371	10	97.63%	97.37%	97.50%	55.171607	328	53	97.35%	97.12%	97.24%	1.927079
		9	370					53	326				
8	77	376	10	97.41%	97.40%	97.41%	55.587305	331	53	97.40%	97.71%	97.55%	1.884540
		10	375					55	332				
9	78	381	11	97.19%	97.18%	97.19%	55.481798	336	54	97.15%	97.46%	97.31%	1.749130
		11	379					55	335				
10	79	385	11	97.22%	97.22%	97.22%	55.030003	341	55	97.51%	97.74%	97.62%	1.726044
		11	384					55	340				
11	80	389	12	97.01%	97.00%	97.00%	54.942187	345	56	97.51%	97.74%	97.62%	1/948337
		12	388					56	344				
12	81	394	12	97.04%	93.03%	97.04%	54.555110	349	57	96.85%	96.30%	96.58%	1.566272
		22	392					57	348				
13	82	397	13	96.83%	96.82%	96.83%	54.143132	353	58	96.84%	96.81%	96.83%	1.589120
		13	396					57	352				
14	83	402	13	96.63%	96.62%	96.63%	54.337591	357	58	96.60%	96.45%	96.52%	1.513309

		14	400					58	356				
15	84	406	15	96.44%	96.42%	96.43%	54.253094	361	59	96.14%	96.01%	96.07%	2.011353
		15	404					60	361				
16	85	410	16	96.24%	96.23%	96.25%	53.734557	366	60	96.47%	96.47%	96.47%	1.459371
		16	408					60	364				
17	86	413	17	96.05%	96.04%	96.04%	53.359145	370	61	96.32%	96.00%	96.16%	1.273876
		17	412					60	368				
18	87	417	18	95.64%	95.85%	95.75%	59.260171	374	62	95.19%	95.85%	95.52%	1.519282
		19	416					62	373				
19	88	421	20	95.46%	95.45%	95.46%	55.870253	378	62	95.02%	95.68%	95.25%	1.367650
		20	429					63	377				
20	89	425	22	95.29%	95.06%	95.17%	56.284392	382	64	95.23%	95.56%	95.39%	1.246092
		21	423					64	381				
21	90	428	24	94.90%	94.67%	94.78%	50.550362	386	64	95.24%	95.33%	95.34%	1.336561
		23	246					65	386				
22	91	431	26	94.31%	94.29%	94.30%	50.081723	390	66	94.30%	94.27%	94.29%	1.076476
		26	429					66	389				
23	92	433	28	93.72%	93.91%	93.82%	54.954328	393	67	93.32%	94.96%	94.13%	0.935527
		29	432					68	393				
24	93	435	31	93.35%	93.32%	93.33%	57.330422	397	69	93.36%	92.04%	92.70%	0.898224
		31	433					69	396				
25	94	436	35	92.57%	92.54%	92.55%	59.419878	400	70	92.75%	91.95%	92.35%	0.792762
		35	434					71	399				
26	95	436	40	91.79%	91.58%	91.85%	56.490077	403	72	92.05%	92.75%	91.90%	0.714929
		39	435					72	402				