



# Accuracy Enhanced Lung Cancer Prognosis for Improving Patient Survivability Using Proposed Gaussian Classifier System

Kaviarasi R<sup>1</sup> · Gandhi Raj R<sup>2</sup>

Received: 30 January 2019 / Accepted: 16 April 2019 / Published online: 24 May 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

Statistical classifier and good accuracy is an essential part of the research in medical data mining. Accurate prediction of lung cancer is an essential step for making effective clinical decisions. After identifying the lung cancer, minimum scopes are available in the medications for patient living in the world. Hemoglobin level and TNM stage wise patients survival period has to be varied. Some group of people survival period is minimal and another group of people survival time is lengthy. This study is aimed to develop a prediction model with new clinical variables to predict lung cancer patients. It's based on revised 8th edition study of TNM in lung cancer. These new attributes are collected from SEER databases, Indian cancer hospitals and research centers. The collected new attributes are classified using supervised machine learning algorithms of linear regression, Naïve Bayes classifier and proposed algorithms of Gaussian K-Base NB classifier. In particular, for TNM stage 1 group of people with normal hemoglobin level (NHBL), that group of lung cancer patient quality of life is highly enhanced. Which proved by using supervised machine learning algorithms. The proposed algorithm classified the database in terms of with respect to tumor size and HB level and the results are confirmed in the R environment. The continuous attribute classification method to prove first level of TNM in lung cancer patient along with standard hemoglobin has to be maintained that the people survivability rate is higher than the smaller level of hemoglobin people survival rate. The Gaussian K-Base NB classifier is more effective than the existing machine learning algorithms for lung cancer prediction model. The proposed classification accuracy has measured using ROC methods.

**Keywords** Lung cancer · Gaussian K-base NB classifier · Hemoglobin level · TNM stage

## Introduction

Lung cancer is mostly affected people by habits and environment pollution. The important two major categories of genes that creation a portion in cancer they are tumor suppressor and oncogenes. Tumor suppressor genes that slow down the cell division, when did not work properly the cell division its can

out of control [1, 2]. When tumor suppressor gene of TP53, RARB, has to be affected by the Human body organ of lung is particularly affected, the lung cell division is uncontrolled. Proto oncogenes help to grow the cell. When to the changes or mutate the bad genes has to be appeared for the human body, this bad genes is called oncogenes [3].

Lung cancer identified by symptoms or health issues of persistent cough, shortness of breath or fatigue. Cancer may be found in the lung tissues, but not the lymph nodes [4, 5]. Lung cancer affected person has been grouped by tumor (T), node (N) and metastasis (M) wise. Lung cancer grouped five stages [6, 7]. They are stage 0, stage I, II, III, IV. The stage 0 identified some symptoms of pain, weight loss and fatigue. Stage I divided into two sub-types of IA and IB. stage I tumor size is less than 3 cm, the tumor has not spread to the lymph node, distant organs or other areas. Stage II divided into two sub-types of IIA and IIB. The stage II Lung cancer tumor size has larger than 7 cm; tumor has spread to the lining of the lung, developed an inflammation and clogs the patient's airway. Stage III divided into two sub-types of IIIA and IIIB.

---

This article is part of the Topical Collection on *Transactional Processing Systems*

---

✉ Kaviarasi R  
kaviarasi.phd@gmail.com

<sup>1</sup> Department of Computer Applications, University College of Engineering (BIT Campus), Anna University, Tiruchirappalli 620024, India

<sup>2</sup> Department of Electrical and Electronics Engineering, University College of Engineering (BIT Campus), Anna University, Tiruchirappalli 620024, India

Stage IV tumor has spread all other parts of the body. The lung cancer patients diagnosed in stage IV people survival rate less than 10% for 5 yr of survival.

In [5, 8, 9] hemoglobin is a protein found in the RBC. The RBC roles are answerable for transport oxygen universally the human frame. A Person has increased hemoglobin by taking iron rich food. Hemoglobin levels are measured by blood test. The normal hemoglobin level is 13.5 g/dl in a man or 12 g/dl in women. In children levels are varied by age. The Low level hemoglobin caused by cancer and then the hemoglobin level is normal than the cancer patient survival time has to be improvised. The stage I group of people with NHBL both men and women has survival time has to be increased. That people survival life has proved that continuous assess to machine learning classification algorithms. Prediction on cancer an early stage it's very helpful for better treatment and then improved to the patient's survival time. The machine learning algorithms of naïve Bayes, regression algorithm, gradient boosting algorithm and new novelty algorithms of GNB algorithm to assess the classification of TNM group and NHBL and LHBL attributes.

## Software introductions

The proposed algorithm implemented using R software which is one of the open sources and it is dominant software in the research environment include. This tool traditional and statistical package such as extensible and several other GUI systems, in addition to the R commander, for interacting with R. Bayesian model are complicated do to solve the mathematical, it has to be feasible to use for practical applications.

## Related work

In the segment we will talk over the current knowledge of lung cancer prophecy and progress of computational methods and to predict lung cancer survivability.

In the past few years, machine learning algorithms are utilized for lung cancer survivability [8, 10]. The clinical data sets of genomic data and tumor suppressor data which are publically not available for the research community, the rest of publically available research data are SEER research database. [10–12] which included tens and thousands of records. A drawback of these datasets is that they do not frequently cover compute genomic and clinical data. Machine learning is basically applied to major categories of classification, clustering, regression and rule extraction. Multi filtration feature selection algorithm used to classify the accuracy of medical data set [13–15].

In [10, 16, 17] supervised machine learning algorithms was using this paper. The supervised model to patient survival prediction system was implemented. In Paper [18] applied Cox regression model to Tumor was measured. In terms was an

appropriate Bayesian network classifier solicitation on being forecast real life medical data set of the size and complication of LUCADA [1, 11, 19, 20]. Lung cancer subsistence estimate retrieved data on SEER database was using ensemble techniques to risk of death after 6 months to 5 year diagnosis is measured, further ensemble elective model to estimating risk of death after 5 years [21–23]. While high white blood cell count patients with Anemia significantly worse radiotherapy response and asses low HB level do not respond as well [12, 24, 25]. Classification of tumor into cancerous and non-cancerous analysis was implemented using NB classifier and C4.5 algorithm the Bayesian is simple, clear accuracy and faster [14, 26].

In paper [6, 27] set of attributes are used to giving accuracy. Three algorithms are applying this work. Compared the accuracy model and finally concluded that Naïve Bayesian network model is giving best accuracy. The prediction model applied naïve Bayes algorithm and NB model. In [28, 29] to assess hemoglobin level to produce quality of life for after identifying an anemia for chemotherapy. HB level is improved to patients drug reaction is changed its very helpful for medications. Clinical stage was assigned after 8th edition of the TNM classification for lung cancer, 736 clinical patients' records is applied for regression analysis is measured NSCLC patients survival time [30, 31].

The proposed model of this study after identifying the lung cancer patients using 8th edition of TNM stage classification used to analyze the tumor, node and metastasis. Real medical data and SEER dataset to classification of low level HB patients with TNM stage 1 and normal HB level patients with TNM stage 1 patient's survival is continuously measured using continuous attribute selection in Gaussian K-base NB classifier to classify the value and ensemble techniques and ROC curve used prove the accuracy of survival classification.

## Materials and methods

In this section, study the materials of lung cancer database, computational method is applied for the supervised machine learning algorithms which we used, then explain to classifications of TNM stage with HB level of lung cancer patients.

## Foundations of data

The clinical data set is collected from various sources of Indian Medical Hospitals, research center and SEER database. SEER is supported by the SRP in NCI's DCCPS. The various sources to lung cancer patient's medical records are analyzed. The lung cancer tumor data is classified in terms of tumor size. (The lung cancer tumor stages are describes the size of the tumor. There are four categories of T1, T2, T3 and T4. The revised 8th edition of TNM data described in the below mentioned Table 1).

**Table 1** As per [2] Revised 8<sup>th</sup> edition of TNM in lung cancer

Tumor Size	Description
TX	TX not has shown a scans. But around powerfulness stand cancer cells present in saliva or in liquefied occupied from the lung.
T0	Not any indication of cancer
Tis	Measured within the internal liner of the lungs
T1	Limited within the lung. T1 is separated into T1a, T1b and T1c giving to the width of the cancer
T1a	1 cm or fewer next to comprehensive portion.
T1b	Cancer is among 1 cm and 2 cm diagonally
T1c	Cancer is among 2 cm and 3 cm diagonally
T2	Cancer is among 3 cm and 5 cm diagonally. It implicates the main airway T2 is separated keen on T2a and T2b
T2a	Cancer among 3 cm too 4 cm
T2b	Cancer among 4 cm too 5 cm
T3	Cancer is among 5 cm to 7 cm. Some more additional tumor in the same lobe of the lung
T4	Greater than 7 cm. Or Certain additional than one lobe of the lung.

**Learning problems and variables**

The data collection is the first step of this study; the real data set is collected from various sources like lung cancer research center, hospitals and SEER databases. A total number of 321 patient’s clinical records are received from the above mentioned various databases.

- a. To precise lung cancer patient’s diagnosed records retrieved from databases.
- b. The next level of this study which stage of people survivability has to improve. Survivability is classified using Naïve Bayes, linear regression and proposed techniques of Gaussian K-base NB classification algorithms.
- c. The classification accuracy is calculated using RMSE method. This method to compare the proposed methods.
- d. The proposed method is producing good accuracy it’s proved by ROC curve.

This proposed method to classify the tumor size and HB level. Then continuous of this study survival time accuracy has to be measured by ROC method. Lung cancer people were classified Normal Hemoglobin Level (NHBL) with TNM

stage1 and Low Hemoglobin Level (LHBL) with TNM stage1. Then ignore the instances of while the cause of death of a patient different from Low HBL or Normal HBL with TNM stage group of lung cancer patient was diagnosed. So, if a patient was diagnosed with a level of HB with TNM stage and improved survivability of lung cancer patient, this patient is considered improved survival either did not improved.

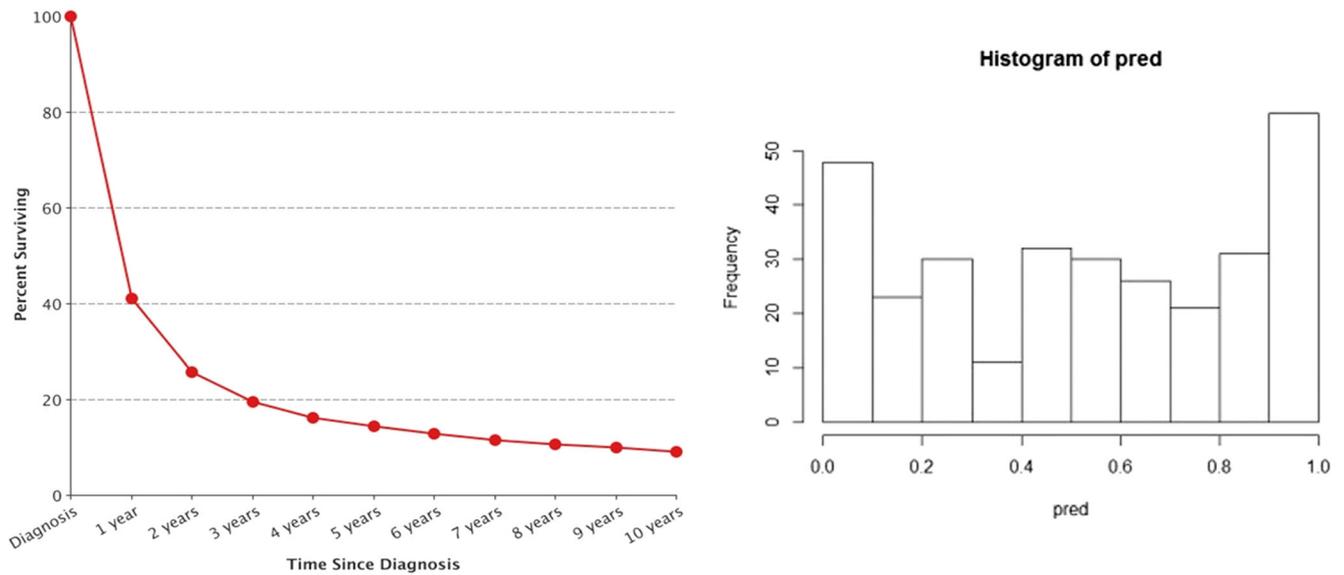
Proposed method of this study lung cancer patients survival prognostic models it’s to use of drug development and essential use feature development of treatments. We used 14 feature attributes to build survivability predictive models. The attributes of the database are described in below mentioned Table 2.

**TNM Stage**

The stage of a lung cancer patient states for people in what way enormous and it has been extent. The stage wise classification was helped for lung cancer patient’s doctors choose which treatment for suitable and also helped for medications for lung cancer patients need. The simple and best suitable one for blood test for common lung cancer patient. It tests to the patient hemoglobin level monitored for some continuous measurement.

**Table 2** Attribute of the database

No	Attributes	Description	Type
1	Age	Age at the time of diagnosis. It’s described young, middle and older.	discrete
2	Gender	Patient’s femininity. Its defined male = 0 and female = 1.	Discrete
3	Radiation sequence with surgery	Sign of whether patient has received radiation	Numeric
4	T	Its telling size of the Tumor	Numeric
5	N	Its telling Lymph Node contribution	Numeric
6	M	Its telling tumor spreading to other organs	Numeric
7	First malignant primary indicator	Based on patient cancer report on clinical and research center database	Numeric
8	Radiation	Suggestion of whether patient has expected radiation.	Numeric
9	Stage	Period of tumor established on T, N and M	Numeric
10	Hemoglobin level	Based on cancer patient report on clinical and research center database.	Numeric
11	WBC	Immune system for lung cancer patients	Numeric
12	RBC	Red blood cell count	Numeric
13	Tumor size	Measurement of tumor size	Numeric
14	Survival time	Amount of months that lung cancer patient are living from date of sympathy.	Discrete



**Fig. 1** Survival period for lung cancer people clinical database (2010–2015) and SEER database to analysis of patient survival time

**Dissemination of existence period**

The dataset retrieved from clinical (hospital) and research center database 321 examples, each of which is a patient record. The complete dissemination of existence period is compared with hemoglobin level is described the following Fig. 1. The Fig. 1 was described for lung cancer people clinical database(2010-2015) and SEER database to analysing of patient survival time.

**Expectation of survival periods of lung cancer patients**

The lung cancer patient training data set has to be preprocessed to select the new attributes of lung cancer people records. The patient’s survival time to assess the patient clinical database in the duration of the year 2010 to 2015.

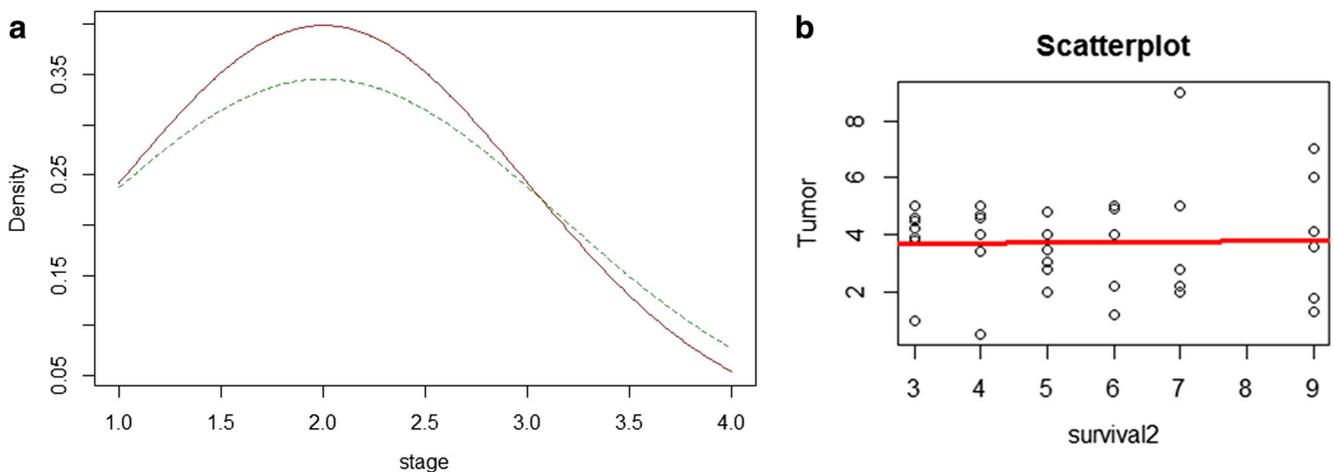
The patient’s survival time has to be varied for based on reports. The reports are based on tumor Size, TNM stage, Hemoglobin

level, gender and age. The survival times are varied based on hemoglobin level with tumor size. The TNM stage I people tumor size is less than 2 mm this value is compared hemoglobin level.

**Analyze survivability through prognosis system**

**Naïve Bayes algorithm**

This for simple prediction algorithm of SML technique it’s used to classify the patient survivability. The above mentioned Table 1, attributes are applied by Naïve Bayes algorithm. In particular, the methods of Naïve Bayes machine learning algorithm to calculate the conditional probabilities. Being naïve Bayes classifier is “independent” first we calculate priori probability of each tag, the probability that it is Hemoglobin Probability is measured and then tumor with HB survivability is classified. Bayesian



**Fig. 2** Classification of Lung cancer patient survival period using machine learning algorithms

theorem is considered conditional “evidence”. Naïve Bayes classifier predicts X values.

The Hemoglobin level is compared to lung cancer patient Tumor, Node and Metastasis value to measure the survival time. Cancer patients HB level high with compared to TNM stage. The TNM stage 1 the cancer people survival time to be increased. It’s proved that the result has to vary by level of HB. The high level of HB has to be improved to their patient’s quality of life.

**Linear regression algorithm**

Linear regression computes the continuous variable it’s used for disease predictions. It’s worked for “independent” or dependent variables. This method to minimize the mean error. Patient survival times measured this method to easy way to speedy appliance for this greater data. The standard glm method in R used gradient boosting ensemble techniques.

**Classification model**

**Naïve Bayes classifier** This NB classifier was used to measure the patient’s survivability. These attributes are conditionally independent not for the dependent attributes. NB classifier measure the probability of survivability. The survivability is possible of this study and more number of features is applied. Let X is a data tuple and H is a hypothesis  $P(HB=Normal | Tumor = stage 1, WBC = 0) = 0.94$ . This NB classifier proved to normal hemoglobin level of peoples with tumor stage 1 has been survival in the world compared to next stage patients.

**Linear regression** Linear regression is a statistical model that used to two continuous variables. Variable X denoted as a predictor and another one Y denoted as a response. These X and Y to determine the strength of the relationship for lung cancer patients. Regression line  $Y = a + bX$ , where X is the explanatory of survival rate and Y is the dependent variables of tumor and hemoglobin levels.

**Results of surviving machine learning algorithms**

a) Naïve Bayes model calculate some dependent variable of tumor stage and hemoglobin probability value. The probability value to classify which stage of people survival has been high. b) Linear regression model calculate some dependent variable Tumor and independent variable survival data. The probability of an analyzed which tumor size having long survival as also presented in the Fig. 2.

**Gaussian K-base NB classifier**

The proposed method based on k attributes. It’s very efficient to classify the data compared to independent of Naïve Bayes classifier and Linear Regression algorithms. This Gaussian referred K attributes and estimating  $P(X_i|C_i)$ . Gaussian K-base Naïve Bayes (GNB) classifier is estimating of  $P(X_i|C_i)$ . Categorical attributes of  $X_k$  states that the significance of attribute  $A_k$  for tuple X. Estimating parameters Y Discrete,  $X_k$  continuous. The attributes are based on K.

---

**Algorithm 1 Classifying the Continuous Significance based on Gaussian distribution**

---

Given X and k values

Maximize  $P(X|C_k)P(C_k)$ , for  $k=1,2,3,4,5$

Step 1: Compute  $P(C_k)$ . The prior probability of each class can be computed based on the training tuples:

Step 2: compute  $P(X|C_k)$  using the following conditional probability.

Step 3: compute  $P(X|C_k)P(C_k)$  for each class

Step 4: A continuous-valued feature is implicit to have a Gaussian distribution through standard deviation  $\sigma$  and mean  $\mu$ .

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

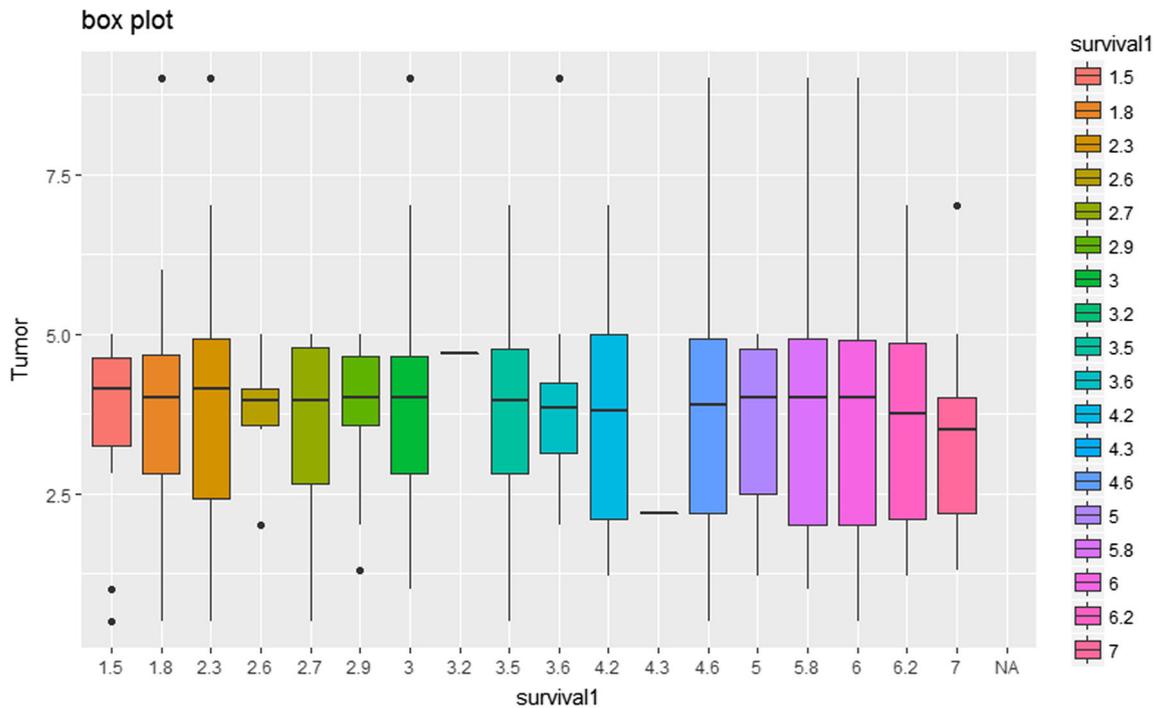
Step 5: The probability  $P(X_k|C_k)$  is given by

$$P(X_k|C_k) = g(X_k, \mu_{C_k}, \sigma_{C_k})$$

Step 6: Estimate  $\mu_{ci}$  and  $\sigma_{ci}$

Step 7: Calculate the mean and standard deviation of attribute  $A_k$  and tuples of class  $C_L$ .

---



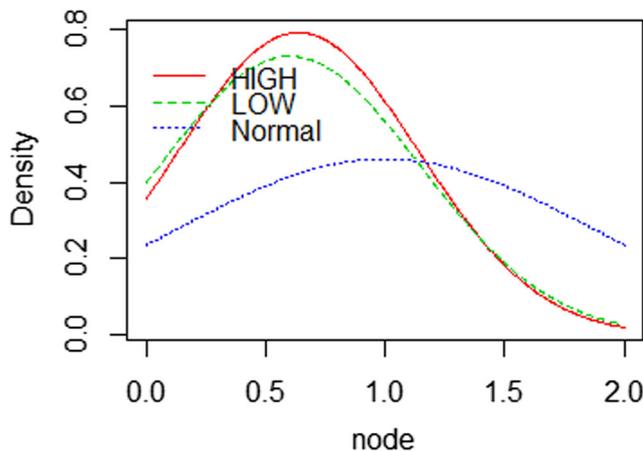
**Fig 3** Survival rate classified as tumor size. The tumor size is 4 levels are categorized. Patient survival rate has been varied T1 stage people survival time has been high, T2 stage people survival rate normal and T3, T4 survival rate has been low

**Performance of Gaussian classification**

Lung cancer prediction system based on after identifying the lung cancer how long duration survived in the world and how to improve patient living time. Patient hemoglobin level varied by men and women. The patient hemoglobin level is normal and also the TNM stage is 1 that type of people survival rate is long otherwise HB level low that people TNM stage is low.

**Hemoglobin level and tumor stage analysis**

This classification classified using Gaussian K-base NB classifier. Figure 3 presents classification results for lung cancer people. Hemoglobin level value and tumor size value used



**Fig. 4** Survival Rate classification for NHBL and LHbL

this classification. These results are less than 2 mm in tumor with NHBL for both men and women survival time has to be high. Tumor size is 2 mm and NHBL for both men and women the people survival time is normal otherwise the survival time is low. Normal human hemoglobin level of male people 13.5 g/dl in a man or less than 12 g/dl in women.

**TNM Stage survival rate analysis**

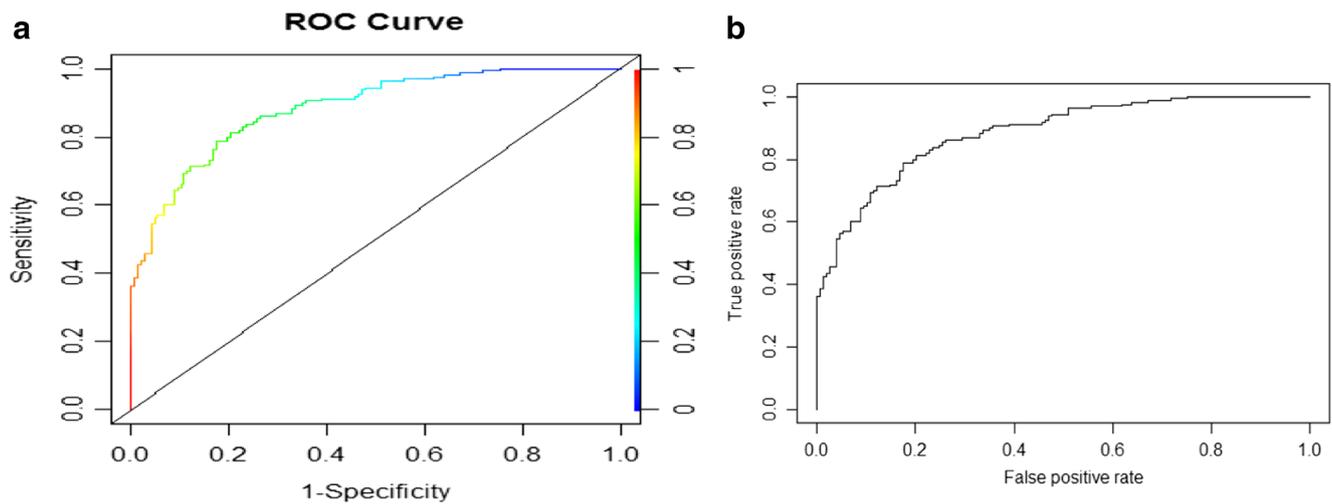
Figure 4 justified Predicted survival time using TNM stage with Hemoglobin levels collective survival rate of Lung cancer patients allowing to their before starting treatment hemoglobin levels. Normal hemoglobin level (NHBL): Male 120–160 g/l, Female 110–150 g/l; low hemoglobin Level (LHbL): Male < 120 g/l and Female ≤ 110 g/l). Complete subsistence of NHbL and LHbL patients lived with disease.

**Classification accuracy**

Lung cancer patient survival accuracy is measured by using ROC methods. The significance test can be used to asses and compared with different classifier. Patient prediction accuracy is one metric of prediction methods of lung cancer patients. The curve plots two parameter values of True positive rate and false positive rate to analyze the accuracy.

True Positive Rate (TPR) = TP/TP + FN

False Positive Rate (FPR) = FP/FP + TN



**Fig. 5** ROC Curve for classification of Gaussian K-Base NB Classifier

The threshold value nearly 0.8 to 1.0 true positive rate has proved that 100% accuracy it's described the below results Fig. 5. The ROC curve is compared to the proposed techniques accuracy. The survival time has to be varied with stage 1 with NHBL people survivability.

Lung cancer people prediction accuracy was performed in terms of ROC for survival time measurement curve it ranks a random positive, higher than a random negative more accurately.

## Results analysis

### Performance evaluation model

Confidence interval also measured by 95% is applied to the pairwise method. Standard error is to be minimum level the significance levels are highly improved. The significance level changed beginning 0 next to the level of 0.08, and then the 95% of confidence interval determination did not comprise 0. The Naive Bayes and linear regression accuracy are compared

**Table 3** Comparison of pairwise ROC method

Linear Regression ~ Naive Bayes Classifier	
Difference between areas	0.0248
Standard Error <sup>a</sup>	0.0940
95% Confidence Interval	-0.159 to 0.209
z statistic	0.264
Significance level	$P = 0.7918$
Linear Regression~ Gaussian K-Base NB Classifier	
Difference between areas	0.259
Standard Error <sup>a</sup>	0.0629
95% Confidence Interval	0.136 to 0.382
z statistic	4.121
Significance level	$P < 0.0001$
Naive Bayes Classifier ~ Gaussian K-Base NB Classifier	
Difference between areas	0.284
Standard Error <sup>a</sup>	0.0758
95% Confidence Interval	0.135 to 0.432
z statistic	3.747
Significance level	$P = 0.0002$

to the proposed Gaussian k-Base techniques. This proposed techniques significance level  $P < 0.0001$  and  $P = 0.0002$ . The ROC curve is also proved 100% of accuracy of new proposed techniques is also mentioned in Tables 3 and 4.

### Comparison of modeling techniques

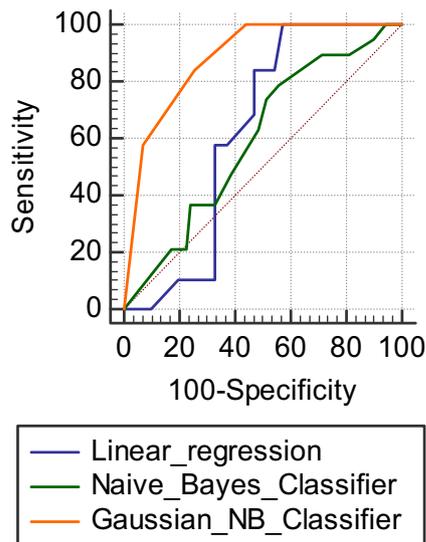
This study to comparison table stated that proposed techniques of Gaussian K-Base NB classifier accuracy compared with existing machine learning algorithms of Naive Bayes Classifier algorithm and linear regression algorithm. The results are showed and described in Fig. 6 Gaussian K-Base NB Classifier accuracy are good for compared to other algorithms.

## Conclusions and discussion

In this paper used different classification techniques with essential Gaussian K-base NB classifier to build a prototypical for survival predictions for lung cancer patients. Prediction accuracies were obtained for continuous survival prediction using classification techniques. The lung cancer patient's prediction to assess survivability using selective attributes. After identifying the lung cancer people to improve the survival time taking healthy foods of rich iron foods, because the hemoglobin level is very helped for patient's quality life. This study is proved that TNM stage 1 people with HB is normal level that type of people survival time is increased is assessed by supervised machine learning algorithms of Gaussian K-

**Table 4** Comparison of proposed method

Method	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
Linear Regression	0.622	0.0569	0.513 to 0.723
Naive Bayes Classifier	0.597	0.0690	0.488 to 0.700
Gaussian K-Base NB classifier	0.881	0.0358	0.795 to 0.940



**Fig. 6** The ROC Curve for proposed techniques of Gaussian K-Base NB Classifier, Naïve Bayes Classifier and Linear Regression

base NB classifier. This accuracy measured using ROC curve. This method to prove accuracy and helpful for doctors drug development for lung cancer patients. Finally improve the medication of patient.

This prediction system is very helpful to the doctors for handling the patients and improves the drug development. In future the patient survival time has to be increased using pre-treatment HB level and pre- diagnosis Hemoglobin level has to survive to improve the patient quality of life.

### Compliance with Ethical Standards

**Conflict of Interest** The Author 1 declares that she has no conflict of interest. Author 2 declares that he has no conflict of interest.

**Ethical Approval** This article does not contain any studies with human participants or animals performed by any of the authors.

### References

- Zhang, J., Gold, K. A., Lin, H. Y., Swisher, S. G., Xing, Y., Lee, J. J., SKim, E., William, Jr., W. N. et al., Relationship Between Tumor Size and Survival in Non-Small-Cell Lung Cancer (NSCLC) An Analysis of the Surveillance, Epidemiology, and End Results (SEER) Registry. *J. Thorac. Oncol.* 10(4), 2015.
- <https://seer.cancer.gov/>
- <https://medlineplus.gov/lungdiseases.html>
- Safiyari, A., Javidan, R. et al., Predicting Lung Cancer Survivability using Ensemble Learning Methods. *Intel. Syst. Conf.*:684–688, 2017.
- Wang, X., Janowczyk, A., Zhou, Y., Thawani, R., Fu, P., Schalper, K., Velcheti, V., Madabhushi, A., et. al. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. *Sci. Rep.*, 2017.
- Esener, I. I. , Ergin, S., Yuksel, T., et. Al. Research Article A New Feature Ensemble with a Multistage Classification Scheme for Breast Cancer Diagnosis. *J. Healthc. Eng.* Volume 2017, Article ID 3895164, 15 pages.
- Blanco, R., and Inza, I., Feature selection in Bayesian classifiers for the prognosis of survival of cirrhotic patients treated with TIPS. *J. Biomed. Inform.* 38:376–378, 2005.
- Chaddad, A., Desrosiers, C., Toews, M., Abdulkarim, B. et al., Predicting survival time of lung cancer patients using radiomic analysis. *Oncotarget* 8(61):104393–104407, 2017.
- Sasikala, S., Appavu, S., and Balamurugan, A., Multi Filtration feature selection (MFFS) to improve Discriminatory ability in clinical data set. *Appl. Comput. Inform.* 12:117–127, 2016.
- Lyncha, C. M., Abdollahib, B., Fuquac, J. D., de Carloc, A. R., Bartholomaic, J. A., Balmann, R. N., van Berkeld, V. H., Frieboesc, H. B., et. Al. Prediction of lung cancer patient survival via supervised machine learning classification techniques, *IJMI* (2017-18).
- Sesen, M. B., Nicholson, A. E., Banares-Alcantara, R., Kadir, T., and Brady, M., Bayesian Networks for Clinical Decision Support in Lung Cancer Care. *PLoS One*, 2013.
- Liu, H.-C., and Huang, M.-Y., The impact of white blood cell count and hemoglobin level on the response to radiotherapy in patients with colorectal cancer. *Biomark. Genom. Med.* 6:171–174, 2014.
- Paulsen, E.-E., Kilvaer, T., Khanekkenari, M. R. et al., CD45RO+ Memory T lymphocytes- a Candidate Marker for TNM Immunoscore in Squamous Non-Small Cell Lung Cancer. *Neoplasia*:839–848, 2015.
- Yeulkar, K., and Sheikh, R., Analysis of SEER Breast Cancer Dataset Using Naive Bayes and C4.5 Algorithm. *IJCST* 8(4), 2017.
- Goldstraw, P., Chansky, K., et. al. The IASLC lung cancer staging project: proposals for revision of the TNM stage grouping in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J. Thor. Oncol.* 2015.
- Port, J. L., Kent, M. S., Korst, R. J., Libby, D., Pasmantier, M., Altorki, N. K., et. al. Tumor Size Predicts Survival Within Stage IA Non-Small Cell Lung Cancer, From the Department of Cardiothoracic Surgery, Weill-Cornell Medical Center, New York, 2003.
- Claribel, P. S., and Filippou, K., et. al. Prognosis in advanced lung cancer – A prospective study examining key clinicopathological factors , Elsevier. *Lung Cancer* 88:304–309, 2017.
- Gerber, D. E., Dahlberg, S. E., Sandler, A. B., Ahn, D. H., Schiller, J. H., Brahmer, J. R., and Johnson, D. H., Baseline tumor measurements predict survival in advanced non-small cell lung cancer. *Br. J. Cancer* 109:1476–1481, 2013.
- Velazquez, E. R., Hober, F., and Aerts, H. J. W. L., et. al. Externally validated HPV-based prognostic nomogram for oropharyngeal carcinoma patients yields more accurate predictions than TNM staging, Elsevier. *Radiother. Oncol.*, 2014.
- Liu, M., Pan, H., Zhang, F., Zhang, Y., Zhang, Y., Xia, H., J Zhu, W. F., and Zhang, X., Identification of TNM stage-specific genes in lung adenocarcinoma by genome-wide expression profiling. *Oncol. Lett.* 6:763–768, 2013.
- Agrawal, A., Misra, S., Narayanan, R., Polepeddi, L., and Choudhary, A., Lung cancer survival prediction using ensemble data mining on SEER data1. *Sci. Program.* 20:29–42, 2012.
- Wu, G., Zhang, D.-Y., Duan, Y.-H., Zhang, Y.-Q., Cui, X.-N., and al Z Luo, e., Correlations of Hemoglobin Level and Perioperative Blood Transfusion with the Prognosis of Gastric Cancer: A Retrospective Study, e-ISSN 1643-3750 ©. *Med. Sci. Monit.* 23: 2470–2478, 2017.
- Derrickerbeck, F. C. et al., The eighth edition TNM stage classification for lung cancer : what does it mean on main street thoracic: feature expert opinion. *Lung Cancer*, 2017.
- Wu, G., Zhang, D.-Y., Duan, Y.-H., Zhang, Y.-Q., Cui, X.-N., and Luo, Z., Correlations of Hemoglobin Level and Perioperative Blood Transfusion with the Prognosis of Gastric Cancer: A

- Retrospective Study. *Clin. Res. Med. Sci. Monit.* 23:2470–2478, 2017.
25. Zhang, J., Gold, K. A., Lin, H. Y., Swisher, S. G., Xing, Y., Lee, J. J., Kim, E. S., William, Jr., W. N. et al., Relationship between tumor size and survival in non-small cell lung cancer (NSCLC): an analysis of the Surveillance, Epidemiology, and End Results (SEER) registry. *J. Thorac. Oncol.* 10(4):682–690, 2015.
  26. Lung cancer staging in the genomics Era, Nguyen, D. M. et al., Thoracic surgery clinics. *Thorac. Surg. Clin.* 16:329–337, 2006.
  27. Langarizadeh, M., and Moghbeli, F., Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review. *Original Paper / ACTA Inform Med.* 24(5):364–369, 2016.
  28. Jean-Loup, Mouysset Beata Freier Joan van den Bosch Charles Briac levaché, Alain, Bols Hans Werner Tessen, laura Belton, g Chet Bohac, Jan-henrik Terwey, Giuseppe Tonini. Hemoglobin levels and quality of life in patients with symptomatic chemotherapy-induced anemia: the eAQUA study. *Cancer Manag. Res.*, 2016.
  29. Yu, K.-H., Zhang, C., Berry, G. J., Altman, R. B., Re, C., Rubin, D. L., Snyder, M. et al., Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.*, 2016.
  30. Y-H Zhang, Y Lu, H Lu, M-W Zhang, Y-M Zhou, X-L Li, P Lv, X-Y Zhao. Pre-treatment hemoglobin levels are an independent prognostic factor in patients with non-small cell lung cancer. *Mol. Clin. Oncol.* 9: 44-49, 201844.
  31. Hosseinzadeh, F., and Joo, A. H. K., et. al. Prediction of lung tumor types based on protein attributes by machine learning algorithms. *Springer Plus* 2:238, 2013.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.