



Machine learning to predict the cancer-specific mortality of patients with primary non-metastatic invasive breast cancer

Cheng-Mao Zhou¹ · Qiong Xue¹ · Ying Wang¹ · Jianhua Tong¹ · Muhuo Ji¹ · Jian-Jun Yang¹

Received: 17 June 2020 / Accepted: 3 September 2020 / Published online: 26 October 2020
© Springer Nature Singapore Pte Ltd. 2020

Abstract

Purpose We used five machine-learning algorithms to predict cancer-specific mortality after surgical resection of primary non-metastatic invasive breast cancer.

Methods This study was a secondary analysis of data for 1661 women with primary non-metastatic invasive breast cancer. The overall patient population was divided into a training group and a test group at a ratio of 8:2 and python was used for machine learning to establish the prognosis model.

Results The machine-learning Gbdt algorithm for cancer-specific death caused by various factors showed the five most important factors, ranked from high to low as follows: the number of regional lymph node metastases, LDH, triglyceride, plasma fibrinogen, and cholesterol. Among the five algorithm models in the test group, the highest accuracy rate was by DecisionTree (0.841), followed by the gbm algorithm (0.838). Among the five algorithms, the AUC values from high to low were GradientBoosting (0.755), gbm (0.755), Logistic (0.733), Forest (0.715), and DecisionTree (0.677).

Conclusion Machine learning can predict cancer-specific mortality after surgery for patients with primary non-metastatic invasive breast.

Keywords Machine learning · Breast cancer · Cancer-specific mortality

Introduction

Breast cancer is the most frequently diagnosed cancer in women. Metastatic breast cancer will develop in approximately 30%–40% of patients with invasive breast cancer [1, 2] and accounts for approximately 15% of cancer deaths of women [3]. Morbidity and mortality tend to increase as the age at diagnosis becomes younger. Studies show a close correlation between a large precancerous space in breast cancer patients and tumor size, histological grade, lymphatic invasion, lymph node metastasis, and prognosis, which could be

used as an important marker to predict prognosis [4, 5]. To explore the risk factors related to the postoperative recurrence of breast cancer, a construct predictive model to help identify patients at high risk of recurrence after surgery, and the timely standardization of treatment for patients with recurrence, will be important to improving the prognosis and quality of life of these patients.

Machine learning is now being applied to medical decisions [6, 7]. The machine-learning processes range from fully manual to fully automated procedures that can process medical data without manual intervention [8]. Machine learning models have produced a means of predicting the risk of in-hospital mortality, using 17 variables to estimate the risk for patients in intensive care unit, with an accuracy of 94% [7]. Deep machine learning has been applied in plastic surgery [9]. However, there are no reported studies on machine learning and cancer-specific survival after invasive breast cancer surgery. The aim of this study is to predict cancer-specific mortality after surgical resection of primary non-metastatic invasive breast cancer using a machine-learning algorithm.

Electronic Supplementary Material The online version of this article (<https://doi.org/10.1007/s00595-020-02170-9>) contains supplementary material, which is available to authorized users.

✉ Cheng-Mao Zhou
zhouchengmao187@foxmail.com

✉ Jian-Jun Yang
yjyangjj@126.com

¹ Department of Anesthesiology, Pain, and Perioperative Medicine, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

Methods

Study population

This study involves a secondary analysis of data from 1661 women with primary non-metastatic invasive breast cancer. The data analysis are available at the BioStudies database [10] (accession number: S-EPMC4658156).

Study variables

Clinical pathology factors, such as axillary lymph node status, menopausal status, age, pathological diagnosis, tumor size, histological grade, hormone receptor and HER-2 status, date of last follow-up, and death related to cancer, were collected. The laboratory data in this study included globulin, total bilirubin (TB), albumin, lactate dehydrogenase (LDH), uric acid, cholesterol, fibrinogen, and triglycerides.

Machine learning

The decision tree (tr) is a binary or multifork tree. This is a supervised learning algorithm.

Logistic regression (lr) is a simple classification algorithm. Logistic regression was used to classify the data by fitting the boundaries of the classification. “Regression” also means the “best fit”, for which it is necessary to find the best fitting parameter. Some optimization methods can be used to calculate the best regression coefficient.

Random Forest (RF) is a classification algorithm that works by forming a large number of decision trees in training and testing. In testing, it outputs classes as class (classification) patterns.

LigthGBM (gbm) is a new member of the boosting set model. The optimizing points of the LightGBM are the “Histogram Algorithm”, “Growth Strategy Optimization of the Tree”, and “High Efficiency”.

Gradient Boosting Decision Tree Classification (Gbdt) is a member of integrated learning. This method adopts the additive model (linear combination of basic functions) and forward step-by-step algorithm. From the weak learning algorithm, a series of weak learners are obtained repeatedly, and then strong learners are obtained by combining weak learners. When each weak learner is a CART tree, it is a lifting tree. A square error is generally used as loss function for regression problems, exponential loss function is used for classification problems, and general loss function is used for general problems.

Statistical analysis

The R language was used for general statistical calculation. The primary endpoint was the cancer-specific survival rate

(CSS) calculated from the date of diagnosis to the date of cancer-related death or last follow-up. The mean and standard deviation were calculated and tested for differences by t-identification. Differences between categories were assessed using the Chi-square test. The python language was used for machine-learning modeling and Pearson’s correlation coefficient was used for correlation analysis. The machine-learning model adopts the following models: logistic regression, decision tree, random, LightGBM, GBDT, and forest. We divided 80% of the overall data into training groups for development and verified 20% by test groups. Missing values were processed by multiple interpolation. The AUC value is between 0 and 1, the greater being better. The related code is shown in Supplementary material 1.

Results

Patients’ characteristics

Table 1 compares the clinical characteristics of the surviving patients (surviving group) and the patients who died (deceased group). The age difference between the surviving group and the deceased group was significant ($p = 0.046$). The tumor size was significantly larger in the deceased group ($p < 0.001$).

Variable importance and correlation analysis

The results of the machine-learning Gbdt algorithm for cancer-specific death caused by various factors rank the top five important factors from high to low as follows: regional lymph node metastasis quantity, LDH, triglyceride, plasma fibrinogen, and cholesterol (Fig. 1). The correlation analysis results show that these five important factors are directly proportional to poor prognosis after breast cancer surgery (Fig. 2).

Machine-learning algorithm models for breast cancer-specific mortality in the training group

Table 2 summarizes the performance characteristics of the models in the training group. Among the five algorithm models, the rate of accuracy was highest for Forest (0.877), followed by the Gradient Boosting algorithm (0.863). Among the five algorithms, the AUC values were 0.973 for Forest, 0.835 for Gradient Boosting, 0.804 for gbm, 0.747 for Logistic, and 0.726 for DecisionTree. Among the five algorithms, Forest had the highest precision rate (0.987) and a recall rate of 0.316 (Fig. 3).

Table 1 Clinical characteristics of the patients

Cancer-specific survival	Alive	Death	<i>p</i> value
<i>N</i>	1365	296	
Age (years)	48.9 ± 10.4	50.3 ± 12.0	0.046
Tumor size	2.7 ± 1.3	3.4 ± 1.8	<0.001
Metastatic lymph node number	2.4 ± 4.6	7.5 ± 8.7	<0.001
Albumin level (g/L, mean ± SD)	43.9 ± 4.0	43.5 ± 4.3	0.020
Globulin level (g/L, mean ± SD)	28.1 ± 4.5	28.9 ± 4.8	0.011
LDH (U/L, mean ± SD)	159.9 ± 57.4	167.0 ± 41.8	0.003
TB (μmol/L, mean ± SD)	13.6 ± 10.1	13.1 ± 4.9	0.833
Uric acid (μmol/L, mean ± SD)	281.3 ± 79.0	290.0 ± 87.3	0.233
Cholesterol (mmol/L, mean ± SD)	5.1 ± 1.1	5.1 ± 1.0	0.253
Triglyceride (mmol/L, mean ± SD)	1.3 ± 0.9	1.5 ± 1.0	0.002
Plasma fibrinogen (g/L, mean ± SD)	2.9 ± 0.7	3.1 ± 0.8	<0.001
Menopause			0.021
Yes	340 (24.9%)	93 (31.4%)	
No	1025 (75.1%)	203 (68.6%)	
ER			<0.001
Negative	472 (34.6%)	142 (48.0%)	
Positive	893 (65.4%)	154 (52.0%)	
PR			<0.001
Negative	381 (27.9%)	118 (39.9%)	
Positive	984 (72.1%)	178 (60.1%)	
HER-2			<0.001
Negative	1076 (78.8%)	192 (64.9%)	
Positive	289 (21.2%)	104 (35.1%)	
Tumor type			0.845
IDC	1326 (97.1%)	289 (97.6%)	
ILC	39 (2.9%)	7 (2.4%)	
Histologic grade			<0.001
G1	448 (32.8%)	114 (38.5%)	
G2	577 (42.3%)	87 (29.4%)	
G3	340 (24.9%)	95 (32.1%)	

SD standard deviation, *IDC* invasive ductal carcinoma, *ILC* invasive lobular carcinoma, *ER* estrogen receptor, *PR* progesterone receptor, *HER-2* human epidermal growth factor receptor-2, *LDH* lactate dehydrogenase, *TB* total bilirubin

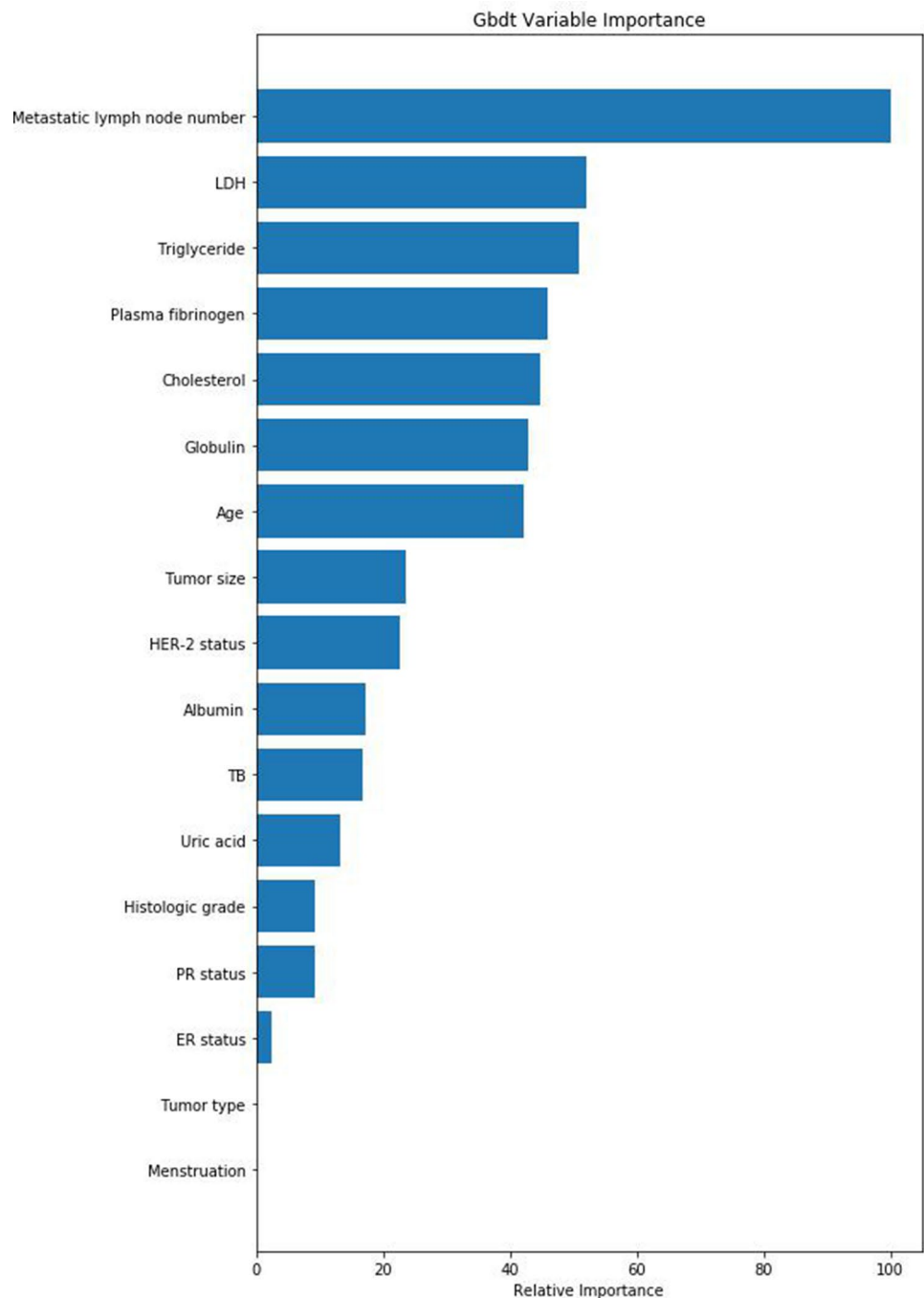
Machine-learning algorithm models for breast cancer-specific mortality in the test group

Table 3 summarizes the performance characteristics of the models in the test group. Among the five algorithm models, DecisionTree had the highest accuracy rate of 0.841, followed by the gbm algorithm (0.838). Among the five algorithms, the AUC values were 0.755 for GradientBoosting, 0.755 for gbm, 0.733 for Logistic, 0.715 for Forest, and 0.677 for DecisionTree. DecisionTree had the highest precision rate (0.667) and gbm had the highest recall rate (0.220) (Fig. 4).

Discussion

Breast cancer is a global issue. According to statistics, about 180,000 cases of breast cancer and 25,000 cases of non-malignant breast tumors are diagnosed each year in the USA [11]. Although the comprehensive treatment of surgery, endocrine therapy, and targeted therapy has improved the overall survival of patients with breast cancer, most patients are still at risk of recurrence and metastasis after surgery. Early postoperative diagnosis and treatment remains a key factor in achieving longer survival. The findings of the present study suggest that machine learning is helpful for

Fig. 1 Variable importance of features included in the machine-learning algorithm for the prediction of postoperative cancer-specific outcomes of patients after breast cancer surgery. Note: *SD* standard deviation, *IDC* invasive ductal carcinoma, *ILC* invasive lobular carcinoma, *ER* estrogen receptor, *PR* progesterone receptor, *HER-2* human epidermal growth factor receptor-2, *LDH* lactate dehydrogenase, *TB* total bilirubin



predicting cancer-specific mortality outcomes after surgery for patients with surgically resected invasive breast cancer. Furthermore, the results of the Gbdt algorithm for cancer-specific mortality showed that the top five important factors were the number of regional lymph node metastases, LDH, triglycerides, plasma fibrinogen, and cholesterol, respectively, in that order.

Axillary lymph node status is an independent predictor of the recurrence and prognosis of breast cancer. The arrangement of axillary lymph node status on the diagnosis

and treatment measures, including the operative method and adjuvant therapy, as well as the psychological status of patients, also has important influence [12]. Previous studies have shown that the 5-year survival rate of breast cancer patients with negative axillary lymph node metastasis exceeds 82%, which decreases to 73% if there are 1–3 lymph node metastases, and then to 46% if there are 4–12 lymph node metastasis [13]. Moreover, about 75% of patients without lymph node metastasis have no relapse within 10 years, whereas 65% of those with 1–3 positive lymph nodes and

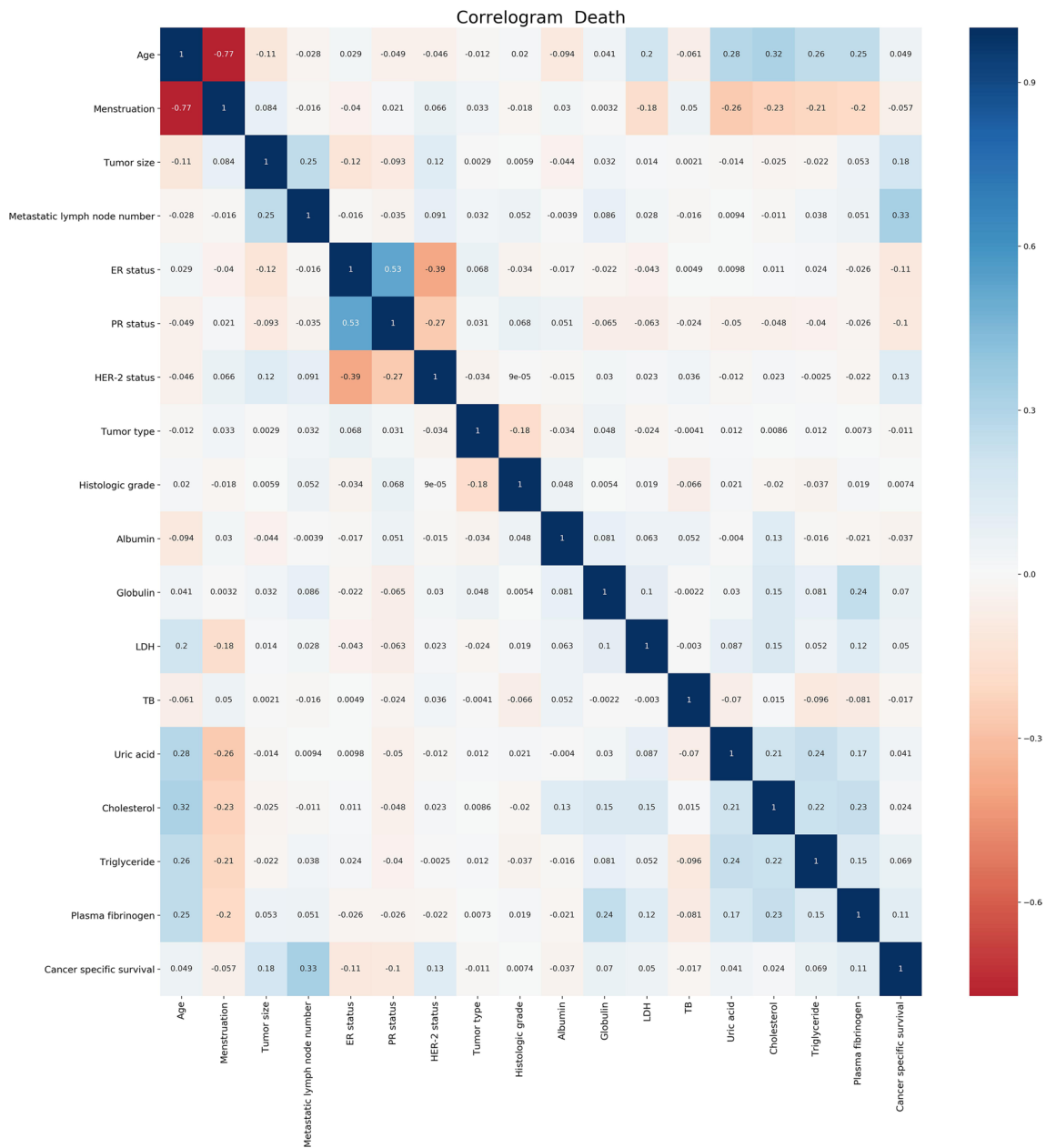


Fig. 2 Correlation among variables

Table 2 Forecast of results for the training group

	Accuracy	Precision	Recall	AUC	F1
Logistic	0.838	0.662	0.190	0.747	30.30
DecisionTree	0.837	0.652	0.181	0.726	0.284
Forest	0.877	0.987	0.316	0.973	0.479
GradientBoosting	0.863	0.982	0.236	0.835	0.381
gbm	0.829	0.917	0.046	0.804	0.088

85% of those with 4 or more positive lymph nodes suffer relapse within 10 years of local treatment [14, 15]. Our research also shows that lymph node metastasis is one of the most important factors for cancer-specific death after breast cancer surgery.

Elevated serum lipid levels may increase the risk of breast cancer and are an important indicator of tumor prognosis. Studies have suggested that the cause of breast cancer is closely related to the concentration of total cholesterol (TC) and triglyceride (TG) in serum [16]. Studies on breast cancer suggest that the onset of breast cancer may be associated

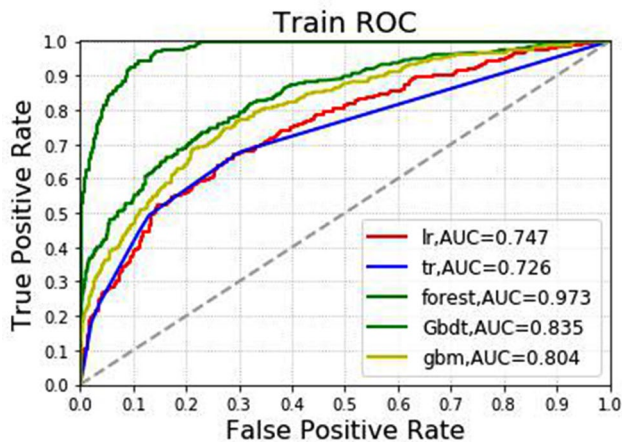


Fig. 3 Different machine-learning algorithms predict the postoperative cancer-specific outcomes of patients after breast cancer surgery in the training group

Table 3 Forecast of results for the testing group

	Accuracy	Precision	Recall	AUC	F1
Logistic	0.835	0.625	0.169	0.733	0.267
DecisionTree	0.841	0.667	0.203	0.677	0.312
Forest	0.817	0.429	0.102	0.715	0.164
GradientBoosting	0.826	0.526	0.169	0.755	0.256
gbm	0.838	0.619	0.220	0.755	0.325

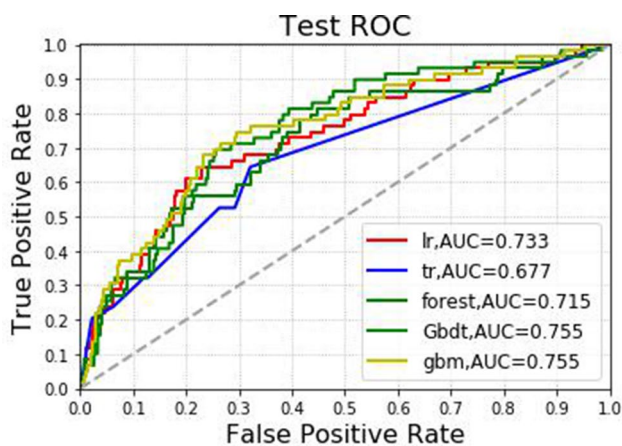


Fig. 4 Different machine-learning algorithms predict the postoperative cancer-specific outcomes of patients after breast cancer surgery in the test group

with hypercholesterolemia [17]. Moreover, high TC, TG, and hypertension are associated with the invasiveness of breast cancer, and these breast cancer patients have higher histologic grades [18]. Studies have also shown that TC and LDL-C levels are significantly higher in breast cancer

patients than in HR+ patients, and that TNBC is often accompanied by low-density lipoprotein cholesterol levels [19]. Hux et al. [20] explained the mechanism of occurrence as follows: in the process of breast cancer proliferation, the metabolism of TG in breast cancer tissue is significantly more active than that in the surrounding tissues, increasing the blood TG level, which decreases the testosterone-estradiol-binding globulin level. This in turn increases the free estradiol concentration in the body, stimulating the abnormal proliferation of mammary epithelium, thus promoting the occurrence and progression of cancer [20]. This explanation is supported by our findings and suggests that high LDH levels are an important risk factor for the survival and prognosis of breast cancer patients [21]. Martina et al. conducted a similar survival analysis of patients with liver metastasis and breast cancer and reported that the overall survival of patients with a high lactate dehydrogenase level was only up to about 10 months, whereas that of those with a normal lactate dehydrogenase level was up to 60 months [22]. A study with more than 10 years of follow-up also showed that the lactate dehydrogenase level was an independent factor for the survival of patients with recurrent breast cancer [23]. Again, these results were similar to those of this study.

Fibrinogen has been associated with distant metastasis of malignant tumors [24]. Altıay et al. identified a trend of higher fibrinogen levels in patients with stage IV lung cancer than in those with stage IIIa/b lung cancer, but the difference did not reach significance [25]. Takuchi et al. also found a strong positive correlation between preoperative fibrinogen levels and the depth of tumor invasion in patients with primary esophageal cancer [26]. Matsuda et al. studied the relationship between changes in fibrinogen levels and prognosis before and after neoadjuvant therapy for esophageal cancer, and found that patients with elevated serum fibrinogen levels had a worse prognosis after neoadjuvant therapy [27]. The results of this study also suggest that serum fibrin is an important risk factor for the prognosis of breast cancer patients.

Albumin, globulin, total bilirubin and urea are all correlated with cancer prognosis. Studies have shown that low serum albumin expression levels are correlated with poor breast cancer prognosis [28]. Additionally, the increase in the C-reactive protein to albumin ratio is an independent risk factor for resectable non-metastatic breast cancer prognosis [29] and serum protein can predict the postoperative survival rates after breast cancer surgery [30]. There are reports that the albumin/globulin ratio can also predict the postoperative prognosis after surgery for small-cell lung cancer. Similarly, the preoperative albumin/globulin ratio can also predict gastric cancer prognosis [31]. Serum albumin and total bilirubin levels are independent risk factors for non-metastatic breast cancer prognosis [32]. Similarly, it has been reported that serum uric acid is an independent risk factor for advanced

gastric cancer prognosis [33, 34]. High serum uric acid is an independent risk factor for breast cancer prognosis [35]. This is also consistent with our research results.

This study was limited by its retrospective nature. First, because the analysis was conducted retrospectively, the data were fixed or unchangeable in the past, which may have led to failure of the subgroup analysis in all dimensions. Second, the predictive machine-learning models may require incremental adjustment, as the predicted effective half-life of historical data for future clinical presentations may be shortened; therefore, more prospective studies are needed. This study does have some over-fitting, although we corrected this in the past by combining regularization and bagging. Furthermore, as this study was a second retrospective analysis, a major defect is that it did not include information about adjuvant treatments and follow-up.

Finally, new support therapy is an important component of breast cancer treatment, which involves giving preoperative systemic treatment to breast cancer patients who are suitable candidates for this, to reduce tumor size and stage. This can provide patients with inoperable breast cancer the opportunity for surgical treatment. It can also allow for breast-conserving surgery in patients who otherwise would not be candidates [36]. Moreover, it can identify tumor drug sensitivity and guide follow-up adjuvant treatment [37]. The treatment plan during follow-up may also be related to breast cancer prognosis. At present, guidelines recommend adjuvant radiotherapy be administered within 6 months after surgery; however, some studies suggest that the earlier the radiotherapy starts, the better the prognosis [38]. Conversely, Caponio et al. [39] analyzed the influence of radiotherapy time on local recurrence and distant metastasis of 615 patients with early breast cancer and found no correlation between postoperative radiotherapy time and local recurrence or distant metastasis risk. Zhang et al. [40] also confirmed that a delay in starting radiotherapy after mastectomy does not increase the risk of local recurrence, distant metastasis, or death.

Conclusion

The findings of our study demonstrate that machine learning can better predict the cancer-specific death outcome of patients undergoing surgical resection for invasive breast cancer. The results of cancer-specific death caused by various factors obtained by the Gbdt algorithm show that the top five important factors are respectively ranked as: regional lymph node metastasis quantity, LDH, triglyceride, plasma fibrinogen, and cholesterol.

Acknowledgements We thank the BioStudies database (public database) for including and providing Professor Xie's original data [41].

Author contributions All authors provided critical feedback and helped shape the research, analysis, and manuscript.

Funding None.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Availability of data and material Data are available at the BioStudies database: <https://www.ebi.ac.uk/biostudies/studies?query=S-EPMC4658156>, accession number: S-EPMC4658156.

Ethics approval and consent to participate This was a secondary analysis using data from the BioStudies database, which is a public database.

References

- Peto R, Davies C, Godwin J, Gray R, Pan HC, Clarke M, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* (London, England). 2012;379:432–44.
- Di LA, Jerusalem G, Petruzella L, Torres R, Bondarenko IN, Khasanov R, et al. Final overall survival: fulvestrant 500 mg vs 250 mg in the randomized CONFIRM trial. *J Natl Cancer Inst*. 2014;106:djt337.
- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA*. 2011;61:69–90.
- Acs G, Paragh G, Chuang S-T, Laronga C, Zhang PJ. The presence of micropapillary features and retraction artifact in core needle biopsy material predicts lymph node metastasis in breast carcinoma. *Am J Surg Pathol*. 2009;33:202–10.
- Shah TS, Kaag M, Raman JD, Chan W, Tran T, Kunchala S, et al. Clinical significance of prominent retraction clefts in invasive urothelial carcinoma. *Hum Pathol*. 2017;61:90–6.
- Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319:1317–8.
- Delahanty RJ, Kaufman D, Jones SS. Development and evaluation of an automated machine learning algorithm for in-hospital mortality risk adjustment among critical care patients. *Crit Care Med*. 2018;46:e481–8.
- Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216–9.
- Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. *IEEE J Biomed health Inf*. 2017;21:4–21.
- Sarkans U, Gostev M, Athar A, Behrangi E, Melnichuk O, Ali A, et al. The BioStudies database—one stop shop for all data supporting a life sciences study. *Nucleic Acids Res*. 2018;46:D1266–70.
- Viale G. Breast cancer. *Lancet*. 2005;365:1727–41.
- Li L, Chen L-Z. Factors influencing axillary lymph node metastasis in invasive breast cancer. *Asian Pac J Cancer Prev*. 2012;13:251–4.
- Fisher B, Bauer M, Wickerham DL, Redmond CK, Fisher ER, Cruz AB, et al. Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. *NSABP Update Cancer*. 1983;52:1551–7.
- Fisher B, Jeong J-H, Anderson S, Bryant J, Fisher ER, Wolmark N. Twenty-five-year follow-up of a randomized trial comparing

- radical mastectomy, total mastectomy, and total mastectomy followed by irradiation. *N Engl J Med.* 2002;347:567–75.
15. Voordeckers M, Vinh-Hung V, Van DSJ, Lamote J, Storme G. The lymph node ratio as prognostic factor in node-positive breast cancer. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* 2004;70:225–30.
 16. Vinnicombe S, Pinto PSM, McCormack VA, Shiel S, Perry N, Dos SSIM. Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data. *Radiology.* 2009;251:347–58.
 17. Hong YC. Serum high-density lipoprotein cholesterol and breast cancer risk by menopause status, body mass index, and hormonal receptor in Korea. *Cancer Epidemiol Biomarkers Prev.* 2009;18:508–15.
 18. Lin X, Hong S, Huang J, Chen Y, Chen Y, Wu Z. Plasma apolipoprotein A1 levels at diagnosis are independent prognostic factors in invasive ductal breast cancer. *Discov Med.* 2017;23:247–58.
 19. Carlson SE. An empirically derived dietary pattern associated with breast cancer risk is validated in a nested case-control cohort from a randomized primary prevention trial. *Clin Nutr Espen.* 2017;17:8–17.
 20. Hux JE. Diabetes mellitus and breast cancer: a retrospective population-based cohort study. *Breast Cancer Res Treat.* 2006;98:349–56.
 21. Cook R. Serum lactate dehydrogenase (LDH) is a significant prognostic variable for survival in patients with metastatic breast cancer—a multivariate analysis. *Eur J Cancer Suppl.* 2008;6:189.
 22. Kamby C, Bruun Rasmussen B, Kristensen B. Prognostic indicators of metastatic bone disease in human breast cancer. *Cancer.* 1991;68:2045–50.
 23. Ryberg M, Nielsen D, Osterlind K, Andersen PK, Skovsgaard T, Dombernowsky P. Predictors of central nervous system metastasis in patients with metastatic breast cancer. A competing risk analysis of 579 patients treated with epirubicin-based chemotherapy. *Breast Cancer Res Treat.* 2005;91:217–25.
 24. Kołodziejczyk J, Ponczek MB. The role of fibrinogen, fibrin and fibrin(ogen) degradation products (FDPs) in tumor progression. *Contemp Oncol.* 2013;17:113–9.
 25. Altıay G, Ciftci A, Demir M, Kocak Z, Sut N, Tabakoglu E, et al. High plasma D-dimer level is associated with decreased survival in patients with lung cancer. *Clin Oncol.* 2007;19:494–8.
 26. Takeuchi H, Ikeuchi S, Kitagawa Y, Shimada A, Oishi T, Isobe Y, et al. Pretreatment plasma fibrinogen level correlates with tumor progression and metastasis in patients with squamous cell carcinoma of the esophagus. *J Gastroenterol Hepatol.* 2007;22:2222–7.
 27. Matsuda S, Takeuchi H, Fukuda K, Nakamura R, Takahashi T, Wada N. Clinical significance of plasma fibrinogen level as a predictive marker for postoperative recurrence of esophageal squamous cell carcinoma in patients receiving neoadjuvant treatment. *Dis Esophagus.* 2014;27:654–61.
 28. Fujii T, Tokuda S, Nakazawa Y, Kurozumi S, Obayashi S, Yajima R, et al. Implications of low serum albumin as a prognostic factor of long-term outcomes in patients with breast cancer. *In Vivo.* 2020;34:2033–6.
 29. Takaaki F, Reina Y, Takahiro T, Toshinaga S, Hiroki M, Satoru Y, et al. Serum albumin and prealbumin do not predict recurrence in patients with breast cancer. *Anticancer Res.* 2014;34:3775–9.
 30. Chung L, Moore K, Phillips L, Boyle FM, Marsh DJ, Baxter RC, et al. Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer. *Breast Cancer Res.* 2014;16:R63.
 31. Zhou T, He X, Fang W, Zhan J, Hong S, Qin T, et al. Pretreatment albumin/globulin ratio predicts the prognosis for small-cell lung cancer. *Medicine (Baltimore).* 2016;95(12):e3097.
 32. Liu X, Meng QH, Ye Y, Hildebrandt MA, Gu J, Wu X. Prognostic significance of pretreatment serum levels of albumin, LDH and total bilirubin in patients with non-metastatic breast cancer. *Carcinogenesis.* 2015;36(2):243–8.
 33. Xue F, Lin F, Yin M, Feng N, Zhang X, Cui YG, et al. Preoperative albumin/globulin ratio is a potential prognosis predicting biomarker in patients with resectable gastric cancer. *Turk J Gastroenterol.* 2017;28(6):439–45.
 34. Yang S, He X, Liu Y, Ding X, Jiang H, Tan Y, et al. Prognostic Significance of serum uric acid and gamma-glutamyltransferase in patients with advanced gastric cancer. *Dis Markers.* 2019;2019:1415421.
 35. Yue C-F, Feng P-N, Yao Z-R, Yu X-G, Lin W-B, Qian Y-M, et al. High serum uric acid concentration predicts poor survival in patients with breast cancer. *Clin Chim Acta.* 2017;473:160–5.
 36. Fisher ER, Wang J, Bryant J, Fisher B, Mamounas E, Wolmark N. Pathobiology of preoperative chemotherapy: findings from the National Surgical Adjuvant Breast and Bowel (NSABP) protocol B-18. *Cancer.* 2002;95(4):681.
 37. Hortobagyi GN. Comprehensive management of locally advanced breast cancer. *Cancer.* 1990;66(Supplement S14):1387.
 38. von MG, Huang C-S, Mano MS, Loibl S, Mamounas EP, Untch M, et al. Trastuzumab emtansine for residual invasive HER2-positive breast cancer. *N Engl J Med.* 2019;380:617–28.
 39. Caponio R, Ciliberti MP, Graziano G, Necchia R, Scognamiglio G, Pascali A, et al. Waiting time for radiation therapy after breast-conserving surgery in early breast cancer: a retrospective analysis of local relapse and distant metastases in 615 patients. *Eur J Med Res.* 2016;21:32.
 40. Zhang W-W, Wu S-G, Sun J-Y, Li F-Y, He Z-Y. Long-term survival effect of the interval between mastectomy and radiotherapy in locally advanced breast cancer]. *Cancer Manag Res.* 2018;10:2047–54.
 41. Wen J, Ye F, Li S, Huang X, Yang L, Xiao X, et al. The Practicability of a Novel Prognostic Index (PI) Model and Comparison with Nottingham Prognostic Index (NPI) in Stages I–III breast cancer patients undergoing surgical treatment. *PLoS ONE.* 2015;10:e0143537.

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