Original Article

Diagnostic accuracy of pleural effusion biomarkers for malignant pleural mesothelioma: a machine learning analysis

Yan Niu[1], Zhi-De Hu[2]

[1]Medical Experiments Center, Inner Mongolia Medical University, Hohhot, China; [2]Department of Laboratory Medicine, the Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China

Contributions: (I) Conception and design: All authors; (II) Administrative support: Y Niu; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: ZD Hu; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Zhi-De Hu, PhD. Department of Laboratory Medicine, the Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China. Email: hzdlj81@163.com.

Background: Some studies have investigated the diagnostic accuracy of pleural effusion (PE) soluble mesothelin-related peptide (SMRP), cytokeratin 19 fragment (CYFRA 21-1), and carcinoembryonic antigen (CEA) for malignant pleural mesothelioma (MPM). However, whether their combination can improve the diagnostic accuracy for MPM remains unclear.

Methods: In this post hoc analysis, 188 subjects, with 27 being diagnosed with MPM, were randomly categorized into training (n=90) and test (n=98) cohorts. We evaluated the diagnostic accuracy of combinational use of PE CEA, SMRP, and CYFRA 21-1 with machine learning approaches, including logistic regression model, linear discriminant analysis (LDA), multivariate adaptive regression splines (MARS), k-nearest neighbor (KNN), gradient boosting machine (GBM), and random forest. Sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC) were used to measure an index test's diagnostic accuracy.

Results: The AUC of the logistic regression model (0.97) was significantly higher than that of CEA (0.75), SMRP (0.86), and CYFRA 21-1 (0.78). The AUCs of MARS, KNN, GBM, and random forest were comparable to that of a single biomarker.

Conclusions: Logistic regression model is a useful machine learning algorithmic approaches to improve the diagnostic accuracy of CEA, SMRP, and CYFRA 21-1. While other machine learning algorithmic strategies (MARS, KNN, GBM, and random forest) cannot improve these biomarkers' diagnostic accuracy.

Keywords: Malignant pleural mesothelioma (MPM); machine learning; diagnosis; soluble mesothelin-related peptide (SMRP); cytokeratin 19 fragment (CYFRA 21-1); carcinoembryonic antigen (CEA)

Received: 31 August 2020; Accepted: 10 September 2020; Published: 30 January 2021.
doi: 10.21037/jlpm-20-90
View this article at: http://dx.doi.org/10.21037/jlpm-20-90

Introduction

Malignant pleural mesothelioma (MPM) is a lethal form of cancer with an inferior prognosis (1). The outcomes of MPM patients can be improved by accurate and timely diagnosis (2). More than half of MPM patients visit the hospital with complaints of chest pain and dyspnea (3). However, these symptoms are not specific to MPM, and MPM's diagnosis is challenging for clinicians. According to the recent guideline released by the British Thoracic Society (BTS), diagnostic pleural aspiration, image-guided cutting needle biopsy, and thoracoscopy are recommended for diagnosing MPM (3). However, these diagnostic tools are invasive and are not available in all centers. Therefore, it is of great value to develop diagnostic tools with reduced

invasiveness. Because pleural effusion (PE) is the most common sign for MPM, soluble biomarkers in PE have been proposed as a tool for diagnostic purposes. During the past decades, several PE biomarkers (4) [e.g., soluble mesothelin-related peptide (SMRP) (5), fibulin-3 (6,7), osteopontin (8,9) and cytokeratin 19 fragment (CYFRA 21-1) (4,10)] have been verified, and their diagnostic accuracy has been evaluated in various studies. However, no biomarker has sufficient diagnostic accuracy for MPM when used alone, according to the guidelines (3,11,12). Therefore, a multiple biomarker approach may be a promising strategy to improve the diagnostic accuracy for MPM.

Machine learning is a type of artificial intelligence. It allows computers to learn with data and build a data model to support a given task with various mathematical and algorithmic approaches (13,14). Machine learning has been used for diagnostic aims in various settings, especially in cancer diagnosis (15). However, the study investigating PE markers' diagnostic accuracy for MPM with machine learning approaches is rare. In this study, we hypothesized that machine learning could improve PE biomarkers' diagnostic accuracy for MPM. We present the following article in accordance with the Standards for Reporting of Diagnostic Accuracy Studies (STARD) reporting checklist (Tables S1,S2) (available at http://dx.doi.org/10.21037/jlpm-20-90).

Methods

Subjects

This is a post hoc analysis of a previous study (16). We obtained the data of this study at the Dryad online repository (http://datadryad.org/review?doi=doi:10.5061/dryad.fg0ft) (17). Briefly, the study mentioned above is a retrospective study performed in a hospital in Japan between September 2014 and August 2016. A total of 240 consecutive patients with undiagnosed PE were enrolled. The diagnostic accuracy of PE SMRP, carcinoembryonic antigen (CEA), and CYRFA 21-1 was assessed using receiver operating characteristic (ROC) curve analysis. Our study excluded the patients with missing values for SMRP, CEA, and CYRFA 21-1. This study was performed with shared data, and we conducted this study following the Declaration of Helsinki (as revised in 2013). Informed consent from the subjects and ethical approval from the authors' institution were waived because the data used in this work are from the internet.

Statistical analysis

In this study, we used machine learning algorithms to evaluate the diagnostic accuracy of PE biomarkers. Briefly, the study cohort was randomly categorized into training and test cohorts. The training cohort was used for model building, and the test cohort was used for validation. The machine learning algorithms used in this study were: logistic regression model (18), linear discriminant analysis (LDA) (19), multivariate adaptive regression splines (MARS) (20), k-nearest neighbor (KNN) (21), support vector machine (SVM) (22), gradient boosting machine (GBM), and random forest. We used the ROC curve to evaluate the diagnostic accuracy of a single biomarker and the machine learning model (23). All analyses were performed with the *caret* package of R (version 4.0.1), and statistical significance was set at $P<0.05$.

Results

Characteristics of the subjects

Figure 1 is a flowchart of the patient selection process. A total of 188 subjects, with 27 being diagnosed with MPM, were included in the present study. They were randomly categorized into training cohort (n=90) and test cohort (n=98). The characteristics of these two cohorts were listed in *Table 1*.

Evaluating the diagnostic accuracy of CEA' SMRP' and CYFRA 21-1 with machine learning algorithms

The diagnostic accuracy of single biomarker and machine learning algorithms was summarized in *Table 2*. When specificity was fixed at 0.94, the sensitivities of CEA, CYFRA 21-1, and SMRP were 0.22, 0.33, and 0.22, respectively. The logistic regression model increased the sensitivity (0.55) without decreasing specificity. Notably, the area under the ROC curve (AUC) of the logistic regression model was higher than that of a single marker and the other machine learning approaches.

Discussion

This study used machine learning approaches to evaluate the diagnostic accuracy of three conventional tumor markers, CEA, SMRP, and CYFRA 21-1, for MPM. The major finding of the present study is that combinational use

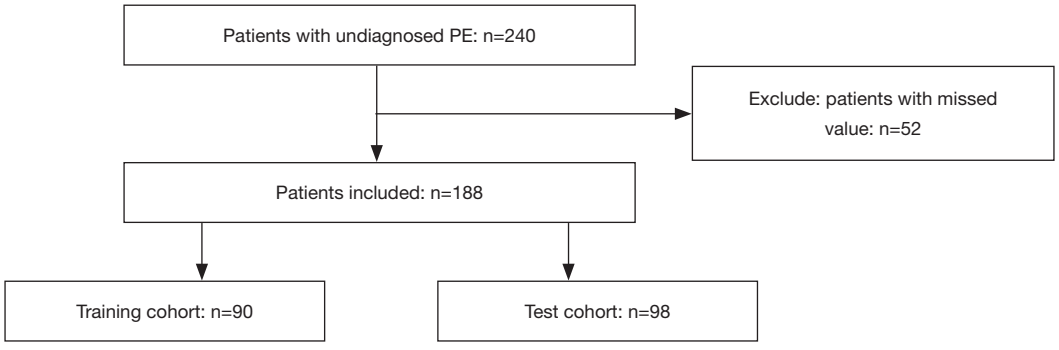

Figure 1 A flowchart of patients selection. PE, pleural effusion.

Table 1 Characteristics of training cohort and test cohort

Characteristics	Training cohort (n=90)	Test cohort (n=98)
Age, years	73 [66–82]	74 [68–82]
CYFRA 21-1 (ng/mL)	32.2 (6.6–112.8)	35.7 (6.5–136.4)
CEA (ng/mL)	1.8 (1.0–3.6)	3.6 (1.3–41.1)
SMRP (nmol/L)	6.3 (3.5–13.8)	5.5 (3.2–9.7)

Data were expressed as median (quartile). SMRP, soluble mesothelin-related peptide; CEA, carcinoembryonic antigen; CYFRA 21-1, cytokeratin 19 fragment.

of these three biomarkers with a logistic regression model can greatly improve the diagnostic accuracy, while other machine learning approaches had limited ability to improve the diagnostic accuracy. Therefore, the logistic regression model represents a potential machine learning algorithm to improve PE tumor markers' diagnostic accuracy for MPM.

This is the second study investigating the diagnostic accuracy of PE tumor markers for MPM, to the best of our knowledge. In the previous study (24), the authors used a logic learning machine (LLM), KNN, artificial neural network (ANN), and decision tree (DT) to evaluate the diagnostic accuracy of PE CEA, SMRP, and CYFRA 21-1 for MPM. They found that the LLM, KNN, ANN, and DT sensitivities were 0.77, 0.56, 0.60, and 0.75, respectively, and the specificities were 0.91, 0.81, 0.86, and 0.92, respectively. Some new machine learning algorithms were used in our study, such as the logistic regression model and random forest. The specificities concluded in our study are generally higher than those of the previous study; however, sensitivities are relatively low. The inconsistency between the present and previous study may be due to the clinical characteristics of MPM and disease profiles of the study cohorts.

Sensitivity and specificity are two primary diagnostic test measures, but their clinical interpretation is not straightforward. The same degree of decrease in sensitivity and specificity can lead to a different number of missed diagnoses and misdiagnoses, which depends on the prevalence of the target disease in the study cohort. By contrast, the positive likelihood ratio (PLR) and negative likelihood ratio (NLR) represent two statistics that are not affected by the target disease's prevalence (25). It is generally accepted that NLR <0.1 or PLR >10 provides strong evidence to rule out or rule in target disease (25). In our study, a PLR of 9.88 was observed in the logistic regression model, suggesting that the positive result of logistic regression is an evidence of MPM. Therefore, the logistic regression model represents a practical algorithm to rule in or out of MPM. AUC is a threshold independent indicator that reflects the overall diagnostic accuracy of an index test. The AUC of the logistic regression model is 0.97, indicating that the logistic regression model is a promising strategy for MPM diagnosis.

This study has two limitations. One limitation is the small sample size in the training and test cohort, and only three markers were considered. The other limitation is the retrospective design, which may introduce patients selection bias. In addition, the results of study have not been validated by other centers. Therefore, further prospective studies with large sample sizes are needed to validate the findings of this study.

Taken together, with three tumor markers in PE, we evaluated the diagnostic accuracy of some machine learning algorithms for MPM. Our results indicate that some machine learning algorithms, such as the logistic regression model, can improve PE tumor markers' diagnostic accuracy.

Table 2 Diagnostic accuracy of single marker and machine learning algorithms

Variables	AUC (95% CI)	Sensitivity	Specificity	PLR	NLR	Accuracy
CEA	0.76 (0.61–0.91)	0.22 (2/9)	0.94 (84/89)	3.96	0.82	0.88 (86/98)
CYFRA 21-1	0.78 (0.59–0.98)	0.33 (3/9)	0.94 (84/89)	5.93	0.71	0.89 (87/98)
SMRP	0.86 (0.76–0.96)	0.22 (2/9)	0.94 (84/89)	3.96	0.82	0.88 (86/98)
Logistic regression model	0.97 (0.94–1.00)	0.55 (5/9)	0.94 (84/89)	9.88	0.47	0.91 (89/98)
LDA	0.89 (0.80–0.98)	0.33 (3/9)	0.94 (84/89)	5.93	0.71	0.49 (48/98)
MARS	0.88 (0.77–0.99)	0.44 (4/9)	0.94 (84/89)	7.91	0.59	0.90 (88/98)
KNN	0.86 (0.72–0.99)	0.33 (3/9)	0.94 (84/89)	5.93	0.71	0.89 (87/98)
Random forest	0.86 (0.72–0.99)	0.44 (4/9)	0.94 (84/89)	7.91	0.59	0.90 (88/98)
GBM	0.84 (0.69–0.99)	0.44 (4/9)	0.94 (84/89)	7.91	0.59	0.90 (88/98)

Numbers in parenthesis indicated the absolute number of patients. AUC, area under the receiver operating characteristic curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio; CEA, carcinoembryonic antigen; CYFRA 21-1, cytokeratin 19 fragment; SMRP, soluble mesothelin-related peptide; LDA, linear discriminant analysis; MARS, multivariate adaptive regression splines; KNN, k-nearest neighbor; GBM, gradient boosting machine.

Give the small sample size and retrospective study design, and future studies are needed to validate our study's findings.

Acknowledgments

Funding: This work was supported by the Natural and Science Foundation of Inner Mongolia Autonomous Region for Distinguished Young Scholars (2020JQ07).

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Laboratory and Precision Medicine*, for the series "Pleural Effusion Analysis". This article has undergone external peer review.

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at http://dx.doi.org/10.21037/jlpm-20-90

Data Sharing Statement: Available at http://dx.doi.org/10.21037/jlpm-20-90

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jlpm-20-90). The series "Pleural Effusion Analysis" was commissioned by the editorial office without any funding or sponsorship. ZDH served as the unpaid Guest Editor of the series and serves as an unpaid executive editor of the *Journal of Laboratory and Precision Medicine* from Nov 2016 to Oct 2021. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was performed with shared data, and the authors conducted this study following the Declaration of Helsinki (as revised in 2013). Informed consent from the subjects and ethical approval from the authors' institution were waived because the data used in this work are from the internet.

References

1. Tsao AS, Wistuba I, Roth JA, et al. Malignant pleural mesothelioma. J Clin Oncol 2009;27:2081-90.

2. Ismail-Khan R, Robinson LA, Williams CC Jr, et al. Malignant pleural mesothelioma: a comprehensive review. Cancer Control 2006;13:255-63.
3. Woolhouse I, Bishop L, Darlison L, et al. British Thoracic Society Guideline for the investigation and management of malignant pleural mesothelioma. Thorax 2018;73:i1-30.
4. Gillezeau CN, van Gerwen M, Ramos J, et al. Biomarkers for malignant pleural mesothelioma: a meta-analysis. Carcinogenesis 2019;40:1320-31.
5. Cui A, Jin XG, Zhai K, et al. Diagnostic values of soluble mesothelin-related peptides for malignant pleural mesothelioma: updated meta-analysis. BMJ Open 2014;4:e004145.
6. Ren R, Yin P, Zhang Y, et al. Diagnostic value of fibulin-3 for malignant pleural mesothelioma: a systematic review and meta-analysis. Oncotarget 2016;7:84851-9.
7. Ledda C, Caltabiano R, Vella F, et al. Fibulin-3 as biomarker of malignant mesothelioma. Biomark Med 2019;13:875-86.
8. Hu ZD, Liu XF, Liu XC, et al. Diagnostic accuracy of osteopontin for malignant pleural mesothelioma: a systematic review and meta-analysis. Clin Chim Acta 2014;433:44-8.
9. Lin H, Shen YC, Long HY, et al. Performance of osteopontin in the diagnosis of malignant pleural mesothelioma: a meta-analysis. Int J Clin Exp Med 2014;7:1289-96.
10. Suzuki H, Hirashima T, Kobayashi M, et al. Cytokeratin 19 fragment/carcinoembryonic antigen ratio in pleural effusion is a useful marker for detecting malignant pleural mesothelioma. Anticancer Res 2010;30:4343-6.
11. Kindler HL, Ismaila N, Armato SG 3rd, et al. Treatment of malignant pleural mesothelioma: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 2018;36:1343-73.
12. Scherpereel A, Opitz I, Berghmans T, et al. ERS/ESTS/EACTS/ESTRO guidelines for the management of malignant pleural mesothelioma. Eur Respir J 2020;55:1900953.
13. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? Clin Chem Lab Med 2018;56:516-24.
14. Gruson D, Helleputte T, Rousseau P, et al. Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. Clin Biochem 2019;69:1-7.
15. Akazawa M, Hashimoto K. Artificial intelligence in ovarian cancer diagnosis. Anticancer Res 2020;40:4795-800.
16. Otoshi T, Kataoka Y, Ikegaki S, et al. Pleural effusion biomarkers and computed tomography findings in diagnosing malignant pleural mesothelioma: a retrospective study in a single center. PLoS One 2017;12:e0185850.
17. Miller GW. Making data accessible: the dryad experience. Toxicol Sci 2016;149:2-3.
18. Zhang Z. Model building strategy for logistic regression: purposeful selection. Ann Transl Med 2016;4:111.
19. Chan YH. Biostatistics 303. Discriminant analysis. Singapore Med J 2005;46:54-61; quiz 62.
20. Friedman JH, Roosen CB. An introduction to multivariate adaptive regression splines. Stat Methods Med Res 1995;4:197-217.
21. Zhang Z. Introduction to machine learning: k-nearest neighbors. Ann Transl Med 2016;4:218.
22. Noble WS. What is a support vector machine? Nat Biotechnol 2006;24:1565-7.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.
24. Parodi S, Filiberti R, Marroni P, et al. Differential diagnosis of pleural mesothelioma using Logic Learning Machine. BMC Bioinformatics 2015;16 Suppl 9:S3.
25. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329:168-9.

doi: 10.21037/jlpm-20-90
Cite this article as: Niu Y, Hu ZD. Diagnostic accuracy of pleural effusion biomarkers for malignant pleural mesothelioma: a machine learning analysis. J Lab Precis Med 2021;6:4.

Supplementary

Table S1 The STARD checklist

Section & topic	No.	Item	Reported on page #
Title or abstract	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	Page 1, line 1
Abstract	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	Page 2, line 18 to 37
Introduction	3	Scientific and clinical background, including the intended use and clinical role of the index test	Page 3, line 41 to 60
	4	Study objectives and hypotheses	Page 3, line 60 to 61
Methods			
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Page 3, line 67 to 68
Participants	6	Eligibility criteria	Page 3, line 68
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	Page 3, line 68
	8	Where and when potentially eligible participants were identified (setting, location and dates)	Page 3, line 68
	9	Whether participants formed a consecutive, random or convenience series	Page 3, line 68
Test methods	10a	Index test, in sufficient detail to allow replication	Page 3, line 69 to 70
	10b	Reference standard, in sufficient detail to allow replication	Page 3, line 65
	11	Rationale for choosing the reference standard (if alternatives exist)	NA
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	NA
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	NA
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	NA
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	NA
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	Page 4, line 77 to 83
	15	How indeterminate index test or reference standard results were handled	NA
	16	How missing data on the index test and reference standard were handled	NA
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	NA
	18	Intended sample size and how it was determined	NA
Results			
Participants	19	Flow of participants, using a diagram	NA
	20	Baseline demographic and clinical characteristics of participants	Table 1
	21a	Distribution of severity of disease in those with the target condition	NA
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	NA
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Table 2
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Table 2
	25	Any adverse events from performing the index test or the reference standard	NA
Discussion	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	Page 6, line 132 to 135
	27	Implications for practice, including the intended use and clinical role of the index test	Page 5, line 122 to 131
Other information	28	Registration number and name of registry	NA
	29	Where the full study protocol can be accessed	NA
	30	Sources of funding and other support; role of funders	Page 6, line 153 to 154

AUC, area under the receiver operating characteristic curve; NA, not applicable.

http://dx.doi.org/10.21037/jlpm-20-90

Table S2 STARD for "Abstracts": essential items for reporting diagnostic accuracy studies in journal or conference abstracts

Section	Item
–	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
Background	Study objectives
Methods	Data collection: whether this was a prospective or retrospective study
	Eligibility criteria for participants and settings where the data were collected
	Whether participants formed a consecutive, random, or convenience series
	Description of the index test and reference standard
Results	Number of participants with and without the target condition included in the analysis
	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
Discussion	General interpretation of the results
	Implications for practice, including the intended use of the index test
Registration	Registration number and name of registry

AUC, area under the receiver operating characteristic curve.

http://dx.doi.org/10.21037/jlpm-20-90