

论著 · 临床研究

无监督学习在临床实验室自动审核规则制定中的实践价值

宋 毅¹,王家驹^{2△},孙良丽³

(1. 重庆市江北区中医院检验科,重庆 400020;2. 四川省达州市中心医院医学检验科,四川达州 635000;
3. 四川省达州市中心医院风湿免疫科,四川达州 635000)

摘要:**目的** 探索自编码机算法对临床实验室自动审核规则制定的临床价值。**方法** 通过机器学习算法中的自编码机算法对四川省达州市中心医院检验科过去 3 年已经人工审核的涉及血清总蛋白(TP)、血清清蛋白(ALB)以及免疫球蛋白的结果进行无监督学习,评估选择合理模型的部分指标,以及对模型分组后的数据进行分布分析,进而探讨模型的合理性。**结果** 通过增加合适的范围内隐层和隐层内神经元的数量,以及采用 dropout 技术会对最终模型的 MSE 带来积极的效果;本文模型认为 TP 介于 44.59~94.56 g/L 之间,ALB 位于 27.14~52.88 g/L 之间,A/G 值位于 0.65~4.86 之间,IgA 位于 0.01~18.74 g/L 之间,IgG 位于 1.34~36.81 g/L 之间,及 IgM 位于 0.001~10.715 g/L 之间的结果可以无需关注,该判断规则与部分实验室已公布的自动审核规则极为接近,本试验数据显示将近 90%判定为无需关注,进而大大提高检验人员对极少数异常数据的关注度。**结论** 无监督学习中的自编码机算法能够在实验室自动审核规则制定中产生积极的价值,但需要进一步对模型产生的规则进行深入解读和分析。

关键词:非监督学习; 自动审核; 临床实验室; 实践价值

DOI:10.3969/j.issn.1673-4130.2018.23.024 **中图法分类号:**R446.1

文章编号:1673-4130(2018)23-2941-05 **文献标识码:**A

The clinical value of unsupervised learning in the establishment of clinical laboratory autoverification rules

SONG Yi¹,WANG Jiasi^{2△},SUN Liangli³

(1. Department of Laboratory Medicine,Chinese Medicine Hospital of Jiangbei District, Chongqing 400020,China. 2. Department of Laboratory Medicine,Dazhou Central Hospital, Dazhou,Sichuan 635000,China. 3. Department of Rheumatism and Immunology, Dazhou Central Hospital,Dazhou,Sichuan 635000,China)

Abstract: Objective To explore the clinical value of autoencoder machine algorithm in the establishment of clinical laboratory autoverification rules. **Methods** Experimental models were generated by deep autoencoder algorithm. Dataset confirmed by manual review was received from clinical laboratory of Dazhou Central Hospital, Sichuan Province. The items observed included serum total protein (TP), albumin (ALB) and immunoglobulin. The distribution characteristics of grouped data were analyzed in order to investigate the validity of the model. **Results** By increasing the number of hidden layers and neurons in the appropriate range, and using dropout technology, the MSE of final model could be positively affected. The rules of the final model, TP=44.59—94.56 g/L, ALB=27.14—52.88 g/L, A/G=0.65—4.86, IgA=0.01—18.74 g/L, IgG=1.34—36.81 g/L and IgM=0.001—10.715 g/L, are very close to the autoverification rules published by some laboratories, and can effectively determine the extreme abnormal data of TP, ALB, A/G, and high level of Ig. This model can determine 90% of the data without concern, and thus greatly improve the attention of inspectors to a small number of abnormal data. **Conclusion** The autoencoder algorithm can generate positive value in the formulation of autoverification rules in laboratory, but further interpretation of the model as well as in-depth analysis is needed.

Key words:unsupervised learning; autoverification; clinical laboratory; practical value

自动审核模块可以加快结果审核以及在一定程度上减少实验室人为错误报告,但是该模块的设置仍然需要各实验室自行设置判读条件,目前尚无统一的自动审核判读规则,而且该判读规则的制定更多依靠

作者简介:宋毅,男,主管技师,主要从事临床生物化学检验的研究。 △ **通信作者,**E-mail:oneclubw@126.com。

本文引用格式:宋毅,王家驹,孙良丽. 无监督学习在临床实验室自动审核规则制定中的实践价值[J]. 国际检验医学杂志,2018,39(23):2941-2945.

于检验人员的临床实际经验。

自编码器算法是非监督学习的重要算法之一^[1], 可以达到主成分分析(PCA)找到原始信息主成分的效果,这得益于机器学习近来发展迅速^[2],该算法对于异常值的检测已得到充分的研究。

在检验人员的日常工作中,免疫球蛋白(Ig)的检测结果的判读常受到多种因素的制约,尤其可能存在的前带现象导致 Ig 测定结果在极高的水平时反而出现正常测定结果的现象^[3],检验人员应该小心防范此类错误结果的发生,结合血清总蛋白(TP)及血清清蛋白(ALB)的测定结果可以减小 Ig 结果错误报告的可能。笔者拟通过自编码深度学习对过往 3 年内人工审核通过的上述检验项目的历史数据进行非监督学习,期望通过本例探讨机器学习在检验结果自动审核规则设置应用中的价值。

1 资料及方法

1.1 数据来源 本次研究所用数据来源于达州市中心医院 2015 年 1 月至 2017 年 12 月间所有同时检测了 TP、ALB 及 Ig(含 IgA、IgG 及 IgM)的 10 759 例患者实际数据,所有数据均已通过人工审核。本研究中舍弃了除标本结果之外的其他所有相关背景资料。所有数据基本资料及对应项目的相关信息见表 1。

1.2 方法 本研究在 R3. 4. 3(x64) for windows 中

进行,无监督学习采用深度自编码神经网络,深度自编码神经网络采用 H2o 机器学习框架(<https://www.h2o.ai/>)。本研究中前后搭建了 7 个模型,对每一个模型进行 10 折交叉验证,根据最小均方误差(MSE)选择相对合适的模型;算法自动将所有数据分为无需关注和需关注两类,无需关注指代模型认为该结果可经由模型推断的规则自动判定为可直接出具报告,需关注指代该结果被模型视为需人工审核后确定是否需额外复查等方可出具报告。笔者将所有原始数据根据所选模型进行标注。因为研究涉及的数据已经经过人工审核,笔者无法确认某个患者的数据由仪器测定时的原始模样,因而无法确定该数据是否经过复查等人工处理手段,也无法正确评估模型的效能。但是,笔者分析了经过训练后的模型标注过的数据的分布范围,尤其关注了被判定为无需关注的数据的分布情况特点,通过对无需关注数据的分布极限与目前实验室采用的相应项目的审核规则进行比较,计算其偏离值(off set),方法为将检测结果值与本实验室采用的该项目参考范围值做了比较,如果在正常参考范围内,即将偏离值(off set)记为 0,如不在正常参考范围内,off set 计算方法为:(测定结果-正常参考范围高值(或低值))/正常参考范围高值(或低值) * 100%,A/G 值未作该计算。

表 1 本研究数据分布及项目检测相关说明表

项目 (单位)	观测组 最低值	观测组 最高值	中位数(P_{25} , P_{75})	测试仪器	试剂厂家	测试方法	线性范围	前带效应
TP(g/L)	28.87	153.52	68.84(63.055~73.71)	BECKMAN AU5821	COULTER Medicalsystem Biotechnology 有限公司	Büret	1~120	—
ALB(g/L)	9.88	57.94	41.46(36.23~45.21)	BECKMAN AU5821	COULTER Medicalsystem Biotechnology 有限公司	Bromocresol green	2~60	—
A/G	0.12	4.86	1.54(1.24~1.82)	BECKMAN AU5821	COULTER —	—(计算值)	—	—
IgA(g/L)	0.01	74.48	2.06(1.51~2.79)	BECKMAN AU5821	COULTER DiaSys Diagnostics Systems GmbH	Immunological transmission turbidimetry	0.08~9.00	>50
IgG(g/L)	0.09	102.76	11.77(9.56~14.52)	BECKMAN AU5821	COULTER DiaSys Diagnostics Systems GmbH	Immunological transmission turbidimetry	0.08~35.00	>80
IgM(g/L)	0.001	42.369	1.329(0.893~1.958)	BECKMAN AU5821	COULTER DiaSys Diagnostics Systems GmbH	Immunological transmission turbidimetry	0.03~8.00	>80

注:—表示未提及或无对应检测仪器及方法等;A/G 值由 ALB/GLB 计算而来,计算公式为:ALB/(TP-ALB)

2 结 果

2.1 不同自编码神经网络 MSE 分布情况 7 种不同自编码神经网络 MSE 分布情况见图 1。

各模型间经单因素方差分析可知, $P>0.05$,组间经 Tukey HSD 检验, $P>0.05$,其中模型 M4(具有 2 个隐层:分别有 200 和 100 个神经元,输入层及隐层 dropout 均为 0.5)的 MSE 具有最小的均值及较小的变异系数(CV),表明模型 M4 在 7 组模型中具有相对最好的稳定性及可靠性,因此,笔者最后选择了模型 M4 做为本次研究的待选模型。

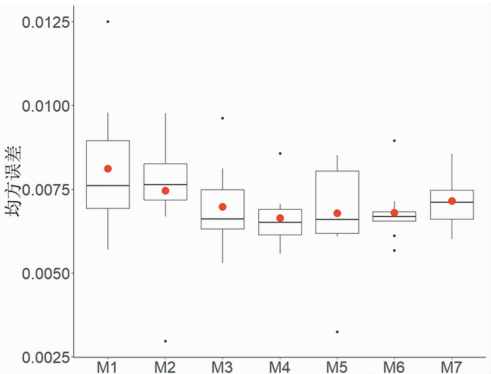
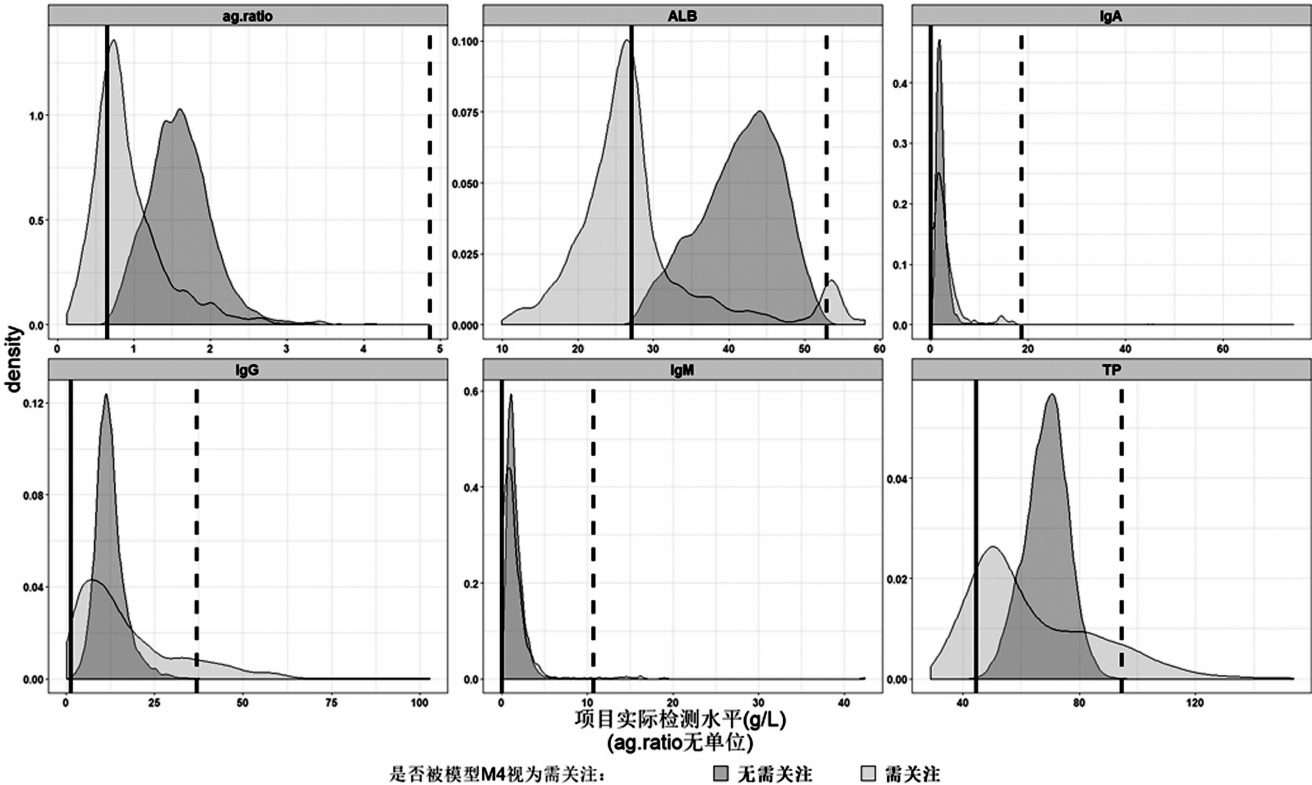


图 1 7 种不同自编码神经网络 MSE 分布情况



注:黑色竖实线表示某项目在无需关注组中的最小判断界限;黑色竖虚线表示某项目在无需关注组的最大判断界限;ag. ratio 表示 A/G 值

图 2 不同标注分组的不同检测项目结果值的分布密度

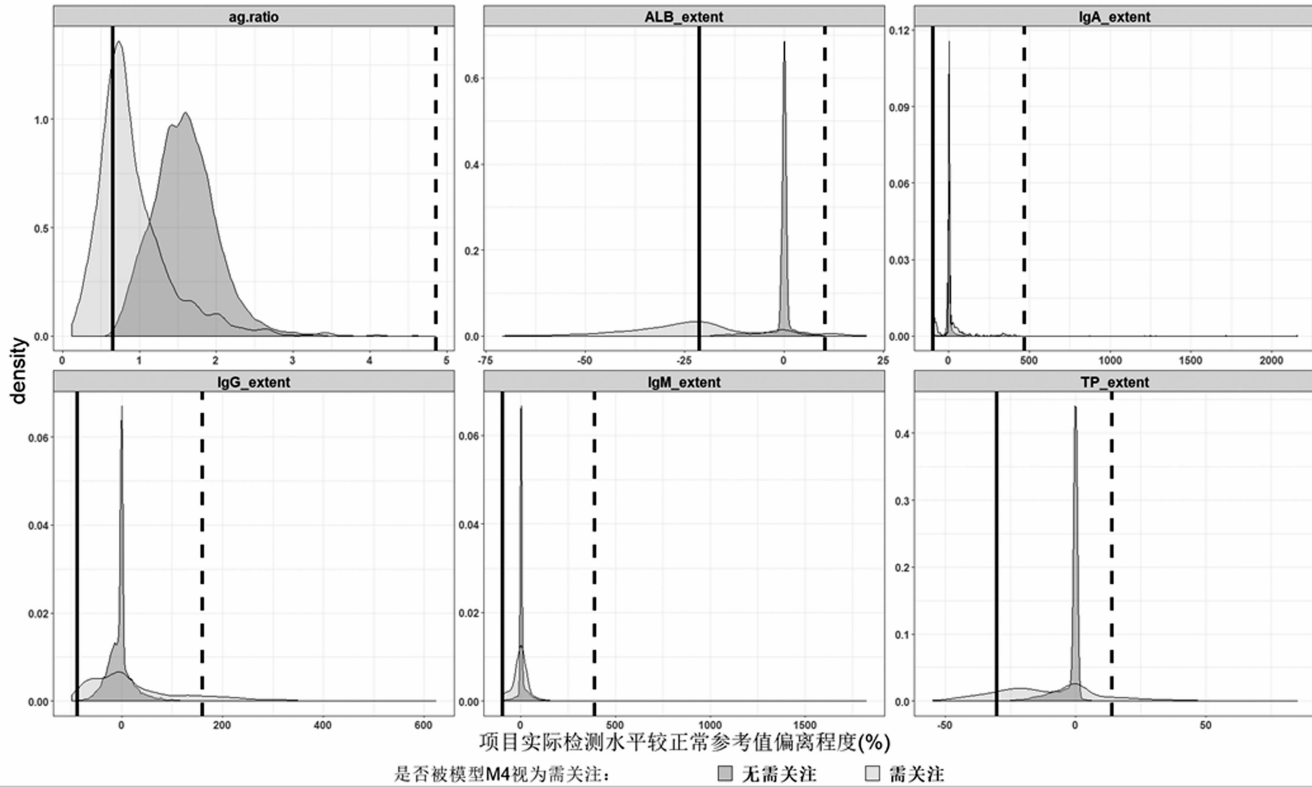


图 3 不同标注分组的不同检测项目结果值较该项目正常参考值偏离程度的分布密度

2.2 不同标注分组的不同检测项目结果值的分布密度 在选择了合适的无监督学习模型后,笔者将该模型在原始数据上进行了标注,分为 2 类:无需关注和需关注。模型 M4 经过了 10 折交叉验证,在每次验证中,分别检出了 988、985、984、1 003、990、982、1 001、

992、999 及 994 例(在所有观察中占比分别为 9.18%、9.16%、9.15%、9.32%、9.20%、9.13%、9.30%、9.22%、9.29%及 9.24%)需关注者。笔者选择了该模型检出率最高者(1 003,9.32%),将该模型标注后的原始结果进行了分布分析,不同标注分组的不同检

测项目结果值的分布密度见图 2。

无需关注组 TP 介于 44.59~94.56 g/L, ALB 位于 27.14~52.88 g/L, A/G 值位于 0.65~4.86, IgA 位于 0.01~18.74 g/L, IgG 位于 1.34~36.81 g/L, 而 IgM 位于 0.001~10.715 g/L, 而需关注组上述各指标分布范围远超前述各范围, 见图 2。

2.3 偏离值结果 对观测指标结果是否偏离正常参考范围而言, 上述项目主要具有以下几个特点: 需关注组观测指标很少出现 off set 为 0, 即该组观测指标检测结果只有少部分处于正常参考范围内, 尤其体现在 TP 和 ALB 上, IgA 除外; TP 及 ALB 在需关注组中其分布均极易集中在出现检测结果偏低时, ALB 尤其如此; 需关注组 Ig 均能有效检出升高者, 需关注组 IgG 升高超过正常参考范围 1.5 倍; 同时, 该组 IgA 和 IgM 则分别超过正常参考范围接近 5 倍及 3 倍。见图 3。

3 讨论

众多检验从业人员一直致力于检验结果自动审核工作的推动及完善。CLSI 于 2006 年发布的关于自动审核的指南明确了实验室检验结果自动审核的定义以及必须服从的规律及建议等^[4]。MATTHEW 等^[5]在一个大型中心的长期应用及验证表明, 服从该指南的自动审核可以大大缩短实验检测周转时间 (TAT), 国内外亦有多家大型医院对自动审核机制的建立以及效能进行了评价^[6-9], 但是笔者也注意到了以下几点: 多家大型中心在部分项目的具体设置上并不一样^[5-11], 如 TP; 在检测结果的判定涉及到多个项目相互关联的时候, 有很多项目并未明确说明其关联规则, 如本研究涉及的 Ig 的结果判断。因此, 寻找多个项目间潜在的关联模式并深入研究以得到更为可靠的审核规则值得尝试。

DEMIRCI 等^[12]第一次尝试了利用人工神经网络 (ANN) 的方法试图建立一个决策模型进而对临床实验室的生化结果进行评估, 其结论表明 ANN 算法在实验室结果审核中可扮演更加重要的角色; 在该文的研究过程中, 更多地利用有监督学习的方式进行了模型的建立。在笔者能查到的文献范围内, 本研究可能是第一次尝试完全排除任何人为干预以及预先标注, 仅靠原始临床检验结果利用无监督学习建立判断机制; 同时, 亦未告诉模型任何关于某检测项目的除测试结果之外的任何其他信息。

首先, 从待选模型的选择来看, 具有 2 个隐层并通过在合适的范围内增加隐层内神经元的数量会对最终模型的 MSE 带来积极的效果, 设置合适的 drop-out 以及 Batch Normalization 可以在一定程度上解决模型的过拟合问题^[13-14], 以提升模型的稳健; 当然, 即使是单个隐层, 通过增加隐层神经元的方法也能够一定程度上提升模型的识别能力。如笔者的待选模型中, M4、M5 和 M6, 均具有较小的 MSE, 但是 M5

因为其较大的 CV 被放弃, M6 较之 M4 具有稍大的平均 MSE 而被舍弃, 同时应看到 M7 模型的稳定性较 M4 稍好, 但其较大的 MSE 使得该模型最终被放弃。在神经网络学习过程中, 促进模型优化的超参数调优是一项相当繁复的工作。

笔者纳入学习的所有数据均已经通过人工审核, 部分数据系复查后所得, 与原始检测数据的比较不可能实现, 这给笔者评估模型的价值带来了一定的困难, 但是, 笔者通过分析无需关注组的数据分布来间接评估模型的价值。

众所周知, TG 由 2 部分组成, 其中之一为 ALB, 其余为球蛋白 (GLB), 而 Ig 是 GLB 的主要组成成分, 其他球蛋白在血液当中的水平较之 Ig 有数量级上的差距; TP 及 ALB 检测方法的成熟度极高, 且有良好的线性范围。MATTHEW 等^[4]的报道中 ALB 即使超过线性范围亦采用实际测定结果。检测结果中如果出现 A/G 极低但 Ig 正常或轻度升高的结果时, Ig 测定结果是否正确是首先值得考虑的问题^[3]。被模型 M4 标注为无需关注的数据组, 其各项检测结果分布见图 2 中所示, 其中本组 TP 及 ALB 的分布界限与夏良裕等^[8]使用的规则界限极其接近 (后者使用范围分别为 45~93 g/L, 27~53 g/L), 但 A/G 则存在较大差距 (后者使用范围为 0.5~2.8), 本研究中将部分 >2.8 的 A/G 值仍然纳入无需关注组, 同时本研究对 A/G 降低有更严格的判断标准, 如果 <0.65 必须纳入需关注组。在 MATTHEW 等^[4]的报道中可见对 Ig 检测结果的审核标准主要参考了试剂线性范围, 而且在检测结果超出线性之后采用实测数据, 没有提及 Ig 结果出现前带现象的相关问题。本研究 Ig 所用试剂均提及了前带现象的存在条件, 因此, 如果在没有 TP 及 ALB 结果作为参考的情况下, 有效识别 Ig 的结果是否存在前带现象将存在一定的困难; 根据笔者的研究结果, 无需关注组 TP 最大可能为 94.56 g/L, 而 A/G 最少可能为 0.65, 此时球蛋白最多可能达到 57.31 g/L, 此为无需关注组 GLB 最大可能浓度, 结合 Ig 试剂出现前带现象的水平, 在本研究中, IgM 和 IgG 无需考虑误判的可能, 而 IgA 在球蛋白中占比超过 87% 时才可能被 M4 模型误判, 显然这是一种极其极端的现象, 事实上本研究中有仅 2 例患者 IgA 超过 50 g/L, 超过 40 g/L 者亦仅 5 例, 且均被标注为需关注, 由此可见, 本研究中 M4 模型均衡考虑了所有观测项目的结果及其相互之间的潜在关系, 比如 Ig 的测定结果总和不应该超过 GLB, 可以在很大程度上保证完全不可能的结果模式被正确识别。本研究的训练数据不够大, 样本丰富程度可能存在一定的缺陷, 应采取继续累加更多的实测数据进入模型建立的训练及迭代中以及超参数的调优等措施来对模型进行进一步完善, 以便避免这种极端情况的发生。图 3 让笔者从偏离正常参考范围的程度大小的角度重新审

视了 M4 模型的价值,同时判断偏离程度在一定程度上避免了部分项目由于年龄、性别等因素所致的正常参考范围的差异带来的潜在影响。

从本研究结论来看,自编码器建立的模型可以将无需关注组的极限范围有效识别,与多家实验室的部分项目的判断范围接近,而且近 90% 的数据组归类为无需关注,进而通过自动审核,这可以让检验人员专注于被标注为需关注的数据的审核。笔者有理由认可无监督学习在检验结果异常结果识别中有其自身的价值。

同时 H2o 框架提供了部署为 JAVA 程序的技术,从而可方便模型部署到实践中去^[15]。但是神经网络的不可解释性让检验从业人员无法从专业的角度去详细解释该模型的细节,作为专业人员,如何寻找能够清晰阐明模型的细节对研究者仍然很重要。

综合本研究结论,笔者认为本研究有以下几点值得关注的地方:(1)无监督学习中的自编码器算法对异常数据识别具有一定的优势,在实践中有一定的价值;无监督学习可以提升研究者对于数据间存在的内部关联的更进一步重视,可能在一定程度上使得检验从业人员可以在自动审核规则的设置上排除潜在的外部因素导致的各种负面影响从而产生积极的作用;(2)模型的识别率及稳定性仍然有进一步提升的空间,更多的数据纳入训练,神经网络隐层和隐层神经元的数量的正确配置以及超参数的调优等都是值得尝试的方法;同时通过模型的迭代应该可以对模型的完善起到积极作用;(3)在检验结果审核过程中,考虑多个项目间的潜在影响是提高审核正确性的必要因素,检验人员应协同临床医生尽量申请相关检测项目,这对保证检验结果的正确是有益的;同时从本文结果可见,在仅仅 5 个项目的相互关联审核中,模型提示的无需关注仅占整体观察数据的 90% 左右,如果纳入更多的指标,需关注占比升高是可以预期的,这虽然在一定程度上可能影响导致 TAT 延长,但是其正面的价值在于尽可能避免了错误结果的误报;受限本次观察项目,以及本研究所用仪器不能提供干扰指数,本研究无法确定无监督学习能否忽略该指数对结果的影响从而正确审核结果,且本研究亦未纳入患者历史数据的比对,更完善的模型需要更多的观测项参与以及对模型的正确理解。

4 结 论

在笔者的检索体系中,本研究是国内首次运用神经网络算法对检验结果的自动审核确认工作做了初步研究,同时更是首次将无监督学习算法引入到该项工作的探索之中。本研究能对实验室配置自动审核规则起到积极的作用,同时促进研究者对自动审核机制的深入认识和进一步的改进;同时,笔者亦希望本

文能引起同行对机器学习算法在检验医学领域的应用的更大兴趣和更深入的探索。

参考文献

- [1] BALDI P. Autoencoders, unsupervised learning, and deep architectures[J]. Int Conf Unsupervised & Transf Learn Workshop, 2011: 37-50.
- [2] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [3] MILLER J J. Interference in immunoassays: avoiding erroneous results[J]. Clin Lab Int, 2004, 28: 14-17.
- [4] Clinical and Laboratory Standard Institute. AUTO10AE: Autoverification of Clinical Lab Test Results; Approved Guideline[R]. Wayne, PA: CLSI, 2006
- [5] KRASOWSKI M D, DAVIS S R, DREES D, et al. Autoverification in a core clinical chemistry laboratory at an academic medical center[J]. J Pathol Inform, 2014, 5(1): 1-13.
- [6] 杨勇, 贾延伟, 曹伟, 等. 生化检验结果自动审核规则的制定及其应用[J]. 临床检验杂志, 2016, 34(2): 152-155.
- [7] 何詠, 余霆, 罗通行. 大量历史数据在生化项目自动审核测试验证中的应用[J]. 中国数字医学, 2017, 12(2): 68-70.
- [8] 夏良裕, 程歆琦, 刘茜, 等. 临床实验室生化免疫项目自动审核程序的建立与应用[J]. 中华医学杂志, 2017, 97(8): 616-621.
- [9] FROM P, SAFFURIELIAS E, BARAK M, et al. Autovalidation rates in an outpatient coagulation laboratory[J]. Int J Lab Hematol, 2015, 37(5): 680-685.
- [10] TORKE N, BORAL L, NGUYEN T, et al. Process improvement and operational efficiency through test result autoverification[J]. Clin Chem, 2005, 51(12): 2406-2408.
- [11] SHIH M, CHANG H, TIEN N, et al. Building and Validating an Autoverification System in the Clinical Chemistry Laboratory[J]. Labmedicine, 2011, 42(11): 668-673.
- [12] DEMIRCI F, AKAN P, KUME T, et al. Artificial neural network approach in laboratory test reporting learning algorithms[J]. Am J Clin Pathol, 2016, 146(2): 227-237.
- [13] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. J Mach Learn Res, 2014, 15(1): 1929-1958.
- [14] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. international conference on machine learning, 2015: 448-456.
- [15] ARNO C, ERIN L, VIRAJ P, et al. Deep learning with h2o[M]. CA: H2o. ai, Inc., 2017.