

基于计算机的木材特征提取和分类识别技术研究综述*

黄鹏桂¹ 赵 璠¹ 李晓平² 关 成² 张严风¹ 吴章康²

(1 西南林业大学大数据与智能工程学院,昆明 650224;

2 西南林业大学国家林业和草原局木材与木竹制品质量检验检测中心,昆明 650224)

摘要:木材由于内部结构和组成成分的差异,使不同种类木材表现出完全不同的理化性质,并决定其不同的用途和商业价格,因此针对木材的分类识别研究具有重要的应用价值。木材分类识别通常经过木材特征提取和基于特征的分类识别这2个步骤。目前木材特征提取主要利用计算机视觉、光谱分析等技术。木材分类识别是基于木材特征的数字化,这一部分可利用计算机算法实现自动识别,较以往人工识别可大幅提高准确度。文中通过分析近20年来木材特征提取和分类识别的相关文献,介绍各种基于计算机的木材特征提取与分类识别技术的特点及适用范围,并结合计算机技术的发展方向探讨木材特征提取与分类识别技术的发展趋势,以期为构建更准确的木材分类识别技术提供参考。

关键词:木材识别,特征提取,特征识别,计算机技术

中图分类号:S781.1

文献标识码:A

文章编号:1001-4241(2020)01-0044-05

DOI:10.13348/j.cnki.sjlyyj.2019.0074.y

Review of Computer-based Wood Feature Extraction and Identification

Huang Penggui¹ Zhao Fan¹ Li Xiaoping²
Guan Cheng² Zhang Yanfeng¹ Wu Zhangkang²

(1 College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming 650224, China;

2 National Forestry and Grassland Administration Center for Quality Supervision and Testing of Wood and Bamboo Products, Southwest Forestry University, Kunming 650224, China)

Abstract: The differences in internal structure and composition bring wood completely different physical and chemical properties, and determine their uses and commercial prices. Therefore, the research on wood identification technology is of important value. As a biomass material, wood has the strong variability in structure and component characteristics, which imposes challenges to the research on wood identification technology. Generally, the wood identification technology can be divided into two steps: feature extraction and feature-based identification. At present, computer vision and spectral analysis are the main technology used to extract wood features. And the technology of wood identification is based on the digitalization of wood feature using the computer algorithms, which can improve the accuracy than manual identification. This paper reviews and analyzes the related literature on the wood feature extraction and identification in the past 20 years, introduces the characteristics and applicable scope of a range of computer based technology for wood feature extraction and classification, and discusses the development trend of the technology of wood feature extraction and classification with the development of computer technology, with the expectation to provide references to build a more accurate wood identification technology.

Keywords: wood identification, feature extraction, feature recognition, computer technology

* 收稿日期:2019-04-07;修回日期:2019-07-02;网络出版日期:2019-08-07。

基金项目:国家自然科学基金项目(31870551);西南林业大学科研启动基金项目(111807)。

第一作者:黄鹏桂,男,硕士生,从事林业信息工程研究,E-mail:1744182303@qq.com。

通信作者:赵璠,男,副教授,硕士生导师,从事林业信息工程研究,E-mail:fzhao@swfu.edu.cn。

木材材料内部结构和组成成分的差异使得不同种类的木材表现出完全不同的理化性质,导致各种木材的用途和商业价格千差万别^[1]。木材一直是国民生产生活中重要和贵重的物质材料,其分类识别技术在各行业应用中显得十分重要。在珍贵木制品的交易过程中,需要通过木材分类识别来防止交易中以次充好的欺诈行为;在海关过境检验中,需要通过木材分类识别来支持违禁物品的检测工作;在建造行业,需要通过木材分类识别来确定木材性质以更充分地了解和利用木材;在医药行业,需要通过药用木材分类识别来确定其药用价值。木材分类识别是通过辨识各种木材在结构特征^[2]和组成成分上的差异来确定木材种类。但木材作为一种生物质材料,同种木材由于生长区域生态环境的不同,结构特征也会有明显不同。因此,准确的木材分类识别技术^[3]一直是国内外研究的热点和难点。本文基于收集整理各种基于计算机的木材特征提取和分类识别技术的特点,总结各种技术的优缺点及适用范围,探讨基于计算机的木材特征提取和木材分类识别技术的发展趋势,以期构建更准确的木材分类识别技术提供借鉴。

1 木材分类识别技术发展历程

木材分类识别技术的发展历程大体可分为3个阶段:1)人工识别阶段。木材分类识别技术最早主要是人工根据木材物理性质的直观表征,通过观察对比逐步确定木材种类。该方法主要依赖识别人员的专业知识和木材的宏观表征^[4],很容易出现误判的情况。2)特征数据库检索阶段。收集整理经过专业识别人员识别确定的木材种类,并将这些识别结果以及对应的宏观和微观结构输入检索数据库,形成后续进行木材分类识别的比对依据^[5]。这个阶段的数据来源依赖于人工判别结果,仍属于人工识别范畴。3)基于计算机的木材自动分类识别阶段。早期的计算机木材自动分类识别技术,主要采用固定式的木材特征提取,然后再借助计算机针对所提取的特征进行分类识别。因此,这种固定特征木材分类识别过程可分为特征提取与分类识别2个步骤,主要研究工作集中在特征提取方式研究,所提取特征对于样本数据的代表性将直接影响分类识别的效果。采用这种分步骤的识别流程的主要原因是早期分类识别算法不够完善,使用未经提取特征的数据做分类识别精度不高^[6]。随着现代计算机技术的发展,现在可借助机器学习^[7]、神经网络模型^[8]等计算机自动学习算法,

在同一模型下同时完成特征提取和分类识别,实现自动化特征提取和识别技术。这样的分类识别技术可避免特征提取和分类识别分步骤进行带来的数据耦合问题,使识别精度进一步提高^[9]。

由于计算机识别时考虑的特征更为全面,且能克服人工误差因素,识别准确率较人工识别大为提高。根据木材特征提取和识别模型的关联,基于计算机的木材自动分类识别技术分为3类:1)固定式特征提取和识别技术;2)自动化特征提取和识别技术;3)语义化特征提取和识别技术。这3类技术存在着连续演变的关系,固定式特征提取和识别技术是现行的主流计算机木材识别方法,主要研究对木材数据特征的提取方式;自动化特征提取和识别技术是计算机分类识别技术在其他分类识别领域取得巨大成功的基础上发展应用到木材识别领域的,这类技术的识别效果由数据驱动;而语义化特征提取和识别技术是目前计算机分类识别技术中尚处在研究阶段的新兴技术,取得了部分成功但还未被应用到木材识别领域,是一种有待深入研究的计算机木材分类识别技术。

2 固定式特征提取和分类识别技术

作为最初始的计算机木材分类识别技术,固定式特征提取与分类识别技术大都是利用计算机视觉技术、光谱分析技术等各种数字图像处理手段^[10-12]以固定映射变化的操作方式从样本中提取有效特征参数,再通过各种统计学或者机器学习方法进行实验对比确定有效的特征参数。这类技术不对木材结构和组分等木材知识进行深入研究,而是以确定木材数据的特征提取方式和构建识别模型作为主要任务,所提取的特征也只是木材种类的间接特征。其识别过程可分为2步:1)样本特征的提取与分析;2)识别模型的搭建与参数调试。由于基于木材在不同角度、层次上提取的特征所建立模型的识别精度不同,特征数据的提取与分析是整个分类识别过程中的主要工作量。该技术主要适用于针对特定木材种类,确定所需识别精度的特征提取方式。

计算机视觉技术^[13]是利用计算机模拟人的视觉功能,从图像中提取有效特征信息。其中,灰度共生矩阵^[14]被认为能有效地描述图片纹理特征^[15],是最为常用的一种计算机视觉技术。采用灰度共生矩阵进行木材纹理图像的分类识别时,是利用灰度共生矩阵提取出的数字统计特征来代替木材纹理特征。白雪冰^[16]、Tou^[17]、Wang^[18]、Pramunendar^[19]都在各自

的实验中采用了灰度共生矩阵进行特征提取。其中Wang^[18]的实验效果最好,针对24种木材的480个样本的识别精度达到了91.7%。除灰度共生矩阵外,还有其他图像纹理特征提取方式。戴天虹等^[20]对RGB图片的3个颜色分量分别提取颜色矩信息,再加上整幅图片的均值和方差,最终构建了11个特征向量对图像进行处理和分级。王亚超^[21]研究了基于小波变换处理木材图片识别的方法,先是用小波处理对图像进行去燥处理,再利用整数5/3小波变换,对木材纹理进行多尺度分解提取特征参数。Sugiarto^[22]利用木材图像的HOG特征进行分类识别。

光谱分析技术是利用不同化合物对光谱吸收率的不同性质来鉴别木材种类,是一种无损或者微损的取样识别方式。常用的光谱识别技术是近红外光谱技术,通过对抽提处理的木粉采集红外光谱,利用傅里叶红外光谱法(FTIR)分析有机高分子化合物的结构和化合组成。胡爱华^[23]、马明宇^[24]、张蓉^[25]、王学顺^[26-27]、Lazarescu^[28]、Nisgoski^[29]针对不同木材系列采用了红外光谱提取木材的间接特征来进行分类识别。Nisgoski^[29]对巴西树种识别实验结果表明,在4 000~10 000 cm⁻¹的光谱范围内,所建立的神经网络没有产生任何识别误差,误差范围为±2%。除红外光谱外,其他光谱也可用于木材特征提取。Lavine^[30]采用拉曼光谱提取特征,对98种温带软木材和硬木以及巴西和洪都拉斯热带木材进行识别。2007年Lebow^[31]用光谱反射法依据木材表面特征对木材进行分类。Piuri^[32]提出了一种基于荧光光谱分析的木材自动分类识别系统。

除上述木材特征选取方式外,还有其他一些特殊的提取方式,如热重曲线^[33-34]、气质联用技术(GC-MS)^[35-37]、应力波^[38]、DNA技术^[39]等。这些特征提取技术的特征提取方式与木材的种类密切相关,不同的特征提取方法带来的识别精度不一致。

3 自动化特征提取和分类识别技术

上述固定特征提取方法需要人为进行特征提取,主要原因是在SVM^[22]和BP神经网络^[40]等识别模型下,以图片像素点作为特征输入模型会使得特征过多,需大量样本数据,同时模型难以训练、预测精度欠缺、计算机资源占用过高等。因此,通过样本数据自动化提取木材样本数据特征的卷积神经网络模型开始替代固定式特征提取与分类识别技术。

自AlexNet^[9]在2012年ImageNet识别大赛夺冠

后,刺激了各种利用卷积神经网络进行图片分类研究的发展。在此后发展出的VGGNet^[41]、GoogLeNet^[42]等网络模型先后以极高的精度夺冠。卷积神经网络由卷积层、池化层、全连接层组成,其中卷积层与池化层起到的作用是自动提取图片的特征,全连接层与BP神经网络的结构是类似的。利用卷积神经网络的识别过程是:首先,直接将图片数据和对应的标签输入到网络模型中,无需指定图片的特征提取方式;然后,利用反向传播算法依据标签和模型的预测值自动调整模型参数,提取适合的图像特征作为分类依据,这降低了对木材特征提取的操作难度,同时避免了人为因素在特征选取过程中的误差。因此,卷积神经网络的发展使木材分类识别技术进入自动化特征提取和分类识别的阶段。但是,卷积神经网络对样本的数量要求较高,且训练模型的运算量远超过固定式特征提取技术。此外,卷积神经网络的识别效果与其深度有很大关系,深度越深,识别效果越好,但也更容易出现过拟合(即对训练数据集的识别精度较高,而对测试数据集的识别精度较低);如果网络深度过浅,则容易出现欠拟合(即对训练数据集和测试数据集的识别精度都较低)。因此,选择合适的网络深度、避免过拟合是构建卷积神经网络的难点。

以卷积神经网络构建一个红木分类识别模型的过程为例,由最新修订的国家红木标准^[43]可知,红木品种间的差异主要集中在横切面、纵切面和轴切面的结构特征上,因此应选用红木的各切面照片和类别标签作为输入,不断进行训练、调整网络结构和模型参数,使得网络模型能够根据现有样本包含的信息,自动学习到能区分各红木类型的特征。最终,训练好的网络模型将能在输入待识别红木切面的照片后,输出各种可能的红木类别以及对应的概率,并将概率值最高的红木类别作为分类识别结果。

需要注意的是,在实际应用中训练样本的数量很难完全达到网络模型的训练要求,故发展出迁移学习^[44]的概念。迁移学习就是直接使用已经利用其他数据集训练好的卷积神经网络作为网络模型,由于这些网络模型已经训练具有图片特征提取能力^[45](即卷积核),故只需去除原先的全连接层,再使用较小样本数据训练出自定义的全连接层,即可达到较高的识别精度。因此,迁移学习降低了对训练样本数量的要求,而识别精度却能达到较高的水平。

4 语义化特征提取和分类识别技术

基于固定式特征提取和自动化特征提取的木材

分类识别,虽然能给出木材分类预测的结果,但却不能直接给出分类依据,无法满足许多木材检测的商业场景需求。2015年提出的RCNN网络^[46]是一种基于深度学习的区域目标检测技术,这种识别模型能预测出图片中多种物体的类别并标注其位置,随后发展出Fast-RCNN网络^[47]、Mask-RCNN网络^[48]等。如果将RCNN网络模型应用到木材识别上,可构建一种基于语义化特征的木材分类识别模型,如图1所示。这种识别模型在输入样本图片后,能框选出并输出木材的结构特征以及木材的种类。根据输出的木材结构特征可以确定木材的种类,同时,根据预测的木材种类结果可以反查木材的结构特征,2种输出结果可以相互交叉验证,保证了结果的正确性。这种交叉验证的识别模型可以有效降低误识别的概率,更为木材分类识别给出了具体的依据。

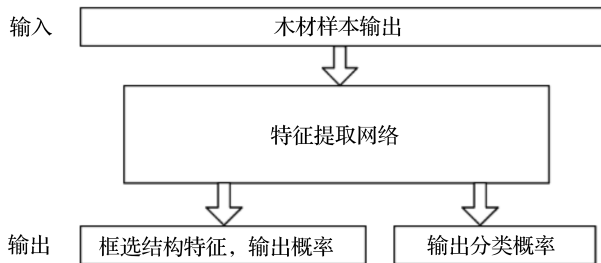


图1 设想中的基于语义化特征的木材分类识别模型

假设以这种技术构建语义化的红木识别模型,在训练时不仅需要输入红木的切面照片和类别标签,还需要标注出每张切面照片所包含的结构特征。这是一个单输入、多输出的模型,模型能够在训练时从带有标签的数据中同时学习到红木的类别信息和对应的结构特征。因此,训练好模型只需要输入红木切片照片后,就能从红木切片的纹理特征中识别出红木类别,并框选输出所用的结构特征。

构建语义化特征的木材识别模型需要大量标注了木材类别并框选了木材结构特征的样本,而相对于RCNN网络模型的整体目标检测识别,木材结构特征相对复杂,难以框选标注。此外,由于这种识别模型结构复杂,训练样本的数据量将远超卷积神经网络,模型训练所需的计算机硬件和运行时间也会成倍增加。因此,语义化特征提取与分类识别技术将是今后木材分类识别技术的重点研究方向。

5 结语及研究展望

随着计算机分类识别技术的发展,木材分类识别

变得更智能化,正在从固定式特征分类识别技术逐步过渡到自动化特征分类识别技术,这种技术因大幅降低了研究难度,将会成为目前木材分类识别的研究热点。然而,基于自动化特征的木材分类识别技术在实际应用中尚存在许多弊端:1)由于依据复杂的函数映射直接给出分类结果,不能根据木材图片中的结构特征显性地给出分类依据;2)这类技术虽然在理论上对限定木材种类内的识别率较高,但是将限定范围外的木材样本数据输入识别模型也会给出限定种类内的错误结果^[40]。

语义化特征提取和分类识别技术可避免上述弊端,但也存在模型构建方法仍有待进一步研究的问题。不过可以肯定的是,基于语义化特征的木材分类识别模型将是未来基于计算机自动木材分类识别的发展方向。这种技术的识别结果将包括分类结果和分类依据,这样就能通过相互验证的方式确保不会有误识别的发生,使计算机自动木材分类识别的应用场景进一步扩大。

参 考 文 献

- [1] 羽杨昆,陈绍志,赵荣,等. 中国主要木材交易市场发展时空变化分析[J]. 世界林业研究,2018,31(3):48-51.
- [2] VINCENT J F. Structure of wood[J]. Current Opinion in Solid State & Materials Science,1998,3(3):228-231.
- [3] FULLER G E. Wood identification[J]. Botanical Gazette,1934,96(2):386-387.
- [4] WHEELER E A, BAAS P. Wood identification - a review[J]. IAWA Journal,1998,19(3):241-264.
- [5] MILLER R B. Wood identification via computer[J]. IAWA Bulletin,1980,1(4):154-160.
- [6] SUYKENS J A K, VANDEWALLE J, MOOR B D. Optimal control by least squares support vector machines[J]. Neural Networks,2001,14(1):23-35.
- [7] MITCHELL T M. Machine learning and data mining[J]. Computer Study,1999,42(2):110-114.
- [8] JAIN A K, MAO J, MOHIUDDIN K. Artificial neural networks: a tutorial[J]. Computer,1996(3):31-44.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems, California, USA,2012.
- [10] PITAS I. Digital image processing algorithms and applications[J]. IEEE Signal Processing Magazine,2002,18(2):58-58.
- [11] 赵景尧,蔡英春. 数字图像相关技术及其在木材科学研究中的应用[J]. 世界林业研究,2015,28(6):53-57.
- [12] HERMANSON J C, WIEDENHOEFT A C. A brief review of machine vision in the context of automated wood identification systems[J]. IAWA Journal,2011,32(2):233-250.

- [13] FAUGERAS O. Three - dimensional computer vision, a geometric viewpoint[J]. *Robotica*,1993,12(5):475 - 475.
- [14] 洪继光. 灰度 - 梯度共生矩阵纹理分析方法[J]. *自动化学报*, 1984,10(1):22 - 25.
- [15] MANJUNATH B S, MA W Y. Texture features for browsing and retrieval of image data[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*,1996,18(8):837 - 842.
- [16] 白雪冰,王克奇,王辉. 基于灰度共生矩阵的木材纹理分类方法的研究[J]. *哈尔滨工业大学学报*,2005,37(12):1667 - 1670.
- [17] TOU J Y, LAU P Y, TAY Y H. Computer vision - based wood recognition system[C]. *International Workshop on Advanced Image Technology*, Kuala Lumpur,2007.
- [18] WANG B H, WANG H J, QI H N. Wood recognition based on grey-level co-occurrence matrix[C]. *International Conference on Computer Application and System Modeling*, Taiyuan, China,2010.
- [19] PRAMUNENDAR R A, SUPRIYANTO C, NOVIANTO D H, et al. A classification method of coconut wood quality based on Gray Level Co-occurrence matrices[C]. *International Conference on Robotics and Automation*, Hong Kong,2013.
- [20] 戴天虹,王克奇,白雪冰,等. 基于神经网络和颜色特征对木材进行分级的分析[J]. *森林工程*,2006,22(1):18 - 20.
- [21] 王亚超. 基于小波变换的木材图像处理技术研究[D]. 呼和浩特:内蒙古农业大学,2013.
- [22] SUGIARTO B, PRAKASA E, WARDOYO R, et al. Wood identification based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) classifier [C]. *Information Systems and Electrical Engineering*, Yogyakarta, Indonesia,2017.
- [23] 胡爱华,邢世岩,巩其亮. 基于 FTIR 的银杏木材鉴别研究[J]. *中国农学通报*,2009,25(4):88 - 92.
- [24] 马明宇,王桂芸,黄安民,等. 人工神经网络结合近红外光谱用于木材树种识别[J]. *光谱学与光谱分析*,2012,32(9):2377 - 2381.
- [25] 张蓉,徐魁梧,张丽沙,等. 基于红外光谱的5种红木树种识别探讨[J]. *林业工程学报*,2014,28(2):95 - 99.
- [26] 王学顺,孙一丹,黄安民. 木材红外光谱的树种识别研究[J]. *森林工程*,2015,31(6):65 - 70.
- [27] 王学顺,孙一丹,黄敏高. 基于 BP 神经网络的木材近红外光谱树种识别[J]. *东北林业大学学报*,2015,43(12):82 - 85.
- [28] LAZARESCU C, HART F, PIROUZ Z, et al. Wood species identification by near - infrared spectroscopy[J]. *International Wood Products Journal*, 2017,8(1):32 - 35.
- [29] NISGOSKI S, OLIVEIRA A A D, MUNIZ G I B D, et al. Artificial neural network and SIMCA classification in some wood discrimination based on near - infrared spectra [J]. *Wood Science*,2017,51(4):929 - 942.
- [30] LAVINE B K, DAVIDSON C, MOORES A J, et al. Raman spectroscopy and genetic algorithms for the classification of wood types[J]. *Applied Spectroscopy*,2001,55(8):960 - 966.
- [31] LEBOW P K, BRUNNER C C, MARISTANY A G, et al. Classification of wood surface features by spectral reflectance [J]. *Wood Fiber Science*,2007,28(1):74 - 90.
- [32] PIURI V, SCOTTI F. Design of an automatic wood types classification system by using fluorescence spectra [J]. *IEEE Transactions on Systems, Man, Cybernetics, Part C*,2010,40(3):358 - 366.
- [33] JAVIER T S, SALVADOR N, MARIO F F, et al. Functional nonparametric classification of wood species from thermal data [J]. *Journal of Thermal Analysis Calorimetry*,2011,104(1):87 - 100.
- [34] MARIO F F, JAVIER T S, ABHIRUP M, et al. A comprehensive classification of wood from thermogravimetric curves [J]. *Chemometrics Intelligent Laboratory Systems*, 2012, 118 (1) : 159 - 172.
- [35] 张洁,夏兆鹏,袁鹏飞,等. 大果紫檀和奥氏黄檀乙醇低毒微损鉴别[J]. *西南林业大学学报*,2015,35(4):75 - 80.
- [36] MANKOWSKI M, BOYD B, HASSAN B, et al. GC - MS Characterizations of termiticidal heartwood extractives from wood species utilized in Pakistan [C]. *The International Research Group On Wood Protection*, Hawaii, USA,2016.
- [37] 张毛毛,蒋劲东,刘波,等. 实时直接分析质谱技术在木材识别研究中的应用[J]. *木材工业*,2019,33(1):29 - 33.
- [38] ROJAS J, ALPUENTE J, POSTIGO D, et al. Wood species identification using stress-wave analysis in the audible range [J]. *Applied Acoustics*, 2011,72(12):934 - 942.
- [39] 王英. 基于 DNA 技术识别樟科木材 [D]. 南京:南京林业大学,2017.
- [40] SADEGHI B H M. A BP - neural network predictor model for plastic injection molding process [J]. *Journal of Materials Processing Technology*,2000,103(3):411 - 416.
- [41] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large - scale image recognition [J/OL]. (2014 - 09 - 04) [2019 - 04 - 07]. <https://arxiv.org/abs/1409.1556>.
- [42] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [J/OL]. (2014 - 09 - 17) [2019 - 04 - 07]. <https://arxiv.org/abs/1409.4842>.
- [43] 殷亚方,姜笑梅,焦立超,等. 新版 GB/T 18107《红木》标准解读 [J]. *木材工业*,2018,32(5):23 - 26.
- [44] PAN S J, YANG Q. A survey on transfer learning [J]. *IEEE Transactions on Knowledge Data Engineering*,2010,22(10):1345 - 1359.
- [45] DAY O, KHOSHGOFTAAR T M. A survey on heterogeneous transfer learning [J]. *Journal of Big Data*,2017,4(1):29.
- [46] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region - based convolutional networks for accurate object detection and segmentation [J]. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 2016,38(1):142 - 158.
- [47] REN S, HE K, GIRSHICH R, et al. Faster R - CNN: towards real - time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39 (6):1137 - 1149.
- [48] HE K, GKIOXARI G, DOLLAR P, et al. Mask R - CNN [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 6 (99):2961 - 2969.