Automated Machine Learning Diagnostic Support System as a Computational Biomarker for Detecting Drug-Induced Liver Injury Patterns in Whole Slide Liver Pathology Images

Munish Puri

Laboratory of Cancer Biology and Genetics, National Cancer Institute, National Institute of Health, Bethesda, Maryland.

ABSTRACT

Drug-induced liver injury (DILI) is a challenging disease to diagnose, a leading cause of acute liver failure, and responsible for drug withdrawal from the market. There is no symptom, no biomarker or test for detection, no therapy, but discontinuation of the drug. Pharmaceutical companies spend huge money, time, and scientific research efforts to test DILI effects and drug efficacy. A preclinical diagnostic support system is designed and proposed for DILI detection and classification on liver biopsy histopathology images. Heterogeneity features and automated machine learning (AutoML) models were tested to classify DILI injury patterns on whole slide image. Fractal and lacunarity values were used to detect hepatocellular necrotic injury patterns caused on a rat liver (in vivo) by 10 drugs at four dose levels. Correlations between fractal and lacunarity values were statistically analyzed for the 10 drugs; the Pearson correlation ($r = 0.9809$), p -value ($1.6612E-06$), and R^2 (0.9582) were found to be high in the case of carbon tetrachloride. The AutoML model was tested to understand the injury patterns on a subset of 1,277 histology images. The AutoML algorithm was able to classify necrotic injury patterns accurately with an average precision of 98.6% on a score threshold of 0.5.

Keywords: drug-induced liver injury, computational biomarker, diagnostic support system, automated machine learning, feature-based DILI detection

INTRODUCTION

DILI is a rare but important disease of the liver, responsible for drug withdrawal from the market in the final phases of clinical trials.¹ Drug-induced liver injury (DILI) remains a significant clinical challenge

in disease management,² a leading cause of acute liver failure in the United States.³ Drug-related hepatotoxicity is uncommon, complex to diagnose, and can be life-threatening. Reported incidences are very less, one in 10,000.⁴ There is no diagnosis, if detected, no effective treatment, other than drug stopping and providing general care.^{5,6} It is challenging to identify DILI affected cases. Thus, hepatotoxicity is the most common reason of post-marketing drug withdrawals.^{7,8}

Liver is the largest organ of the body, placed in the right upper quadrant of the abdomen, and is responsible for detoxification of drugs, toxins, and other substances. Liver function is governed by hepatocytes, the basic architectural cells of the liver. Drug metabolism occurs in the liver, and most of the drug-associated injuries result from drug metabolites. DILI is often nonspecific and misdiagnosed with other liver disease patterns, which affects patient care. In routine clinical practice, pathologists visually examine liver biopsy under a microscope, confirm the injury patterns, and perform causality assessment. Histology features of injury patterns, differential diagnosis, and expert analysis are the integrated biopsy assessments.⁸ A diagnosis of DILI in early stages is important but a complex and challenging task. Accurate evaluation and causality assessment are the key aspects of early detection. Till date, there is no scoring system available for DILI injury assessment.⁹ The link between drug dose intake and injury patterns can be helpful to understand causality assessment. For clinical feature assessment, Roussel Uclaf Causality Assessment Method (RUCAM), endorsed by the Council of International Organizations of Medical Sciences,¹⁰ and DILI scale¹¹ can be used in practice. Also, chronic DILI can be progressive toward liver fibrosis and cirrhosis, if the drug is not discontinued for >3 months.¹² Pharmaceutical companies spend huge amounts of money on clinical trials and drug discovery.¹³ Drug development is a highly complex and multidimensional process; involves very long time, high cost, and huge scientific efforts; and takes 10–15 years for a drug to hit the market, after satisfying strict norms of the Food and Drug Administration approval process.¹⁴

To understand the phenotype of DILI, LiverTox, a collaborative effort by the National Institute of Diabetes and Digestive and Kidney Diseases and the National Library of Medicine of the National Institutes of Health (NIH, USA), provides open-source updated information on hepatotoxicity and DILI. According to LiverTox, DILI can be categorized in 12 phenotypes or overall clinical characteristics based on symptoms.¹⁵ Acute hepatic necrosis is one of the 12 phenotypes of DILI. Drug-related hepatotoxicity can be further subclassified into intrinsic and idiosyncratic. Intrinsic hepatotoxicity is predictable and mostly dose-dependent and reversible. Acetaminophen and isoniazid are typical examples of drugs causing dose-dependent injuries, when consumed over the recommended doses.¹⁶ A pattern of acute hepatic necrotic injury can be seen in 50% of hepatotoxin-induced injuries,¹⁷ which develops when most of the hepatic liver cells produce necrosis.

In routine clinical practice, pathologists visually examine liver biopsy sections under a microscope to assess DILI injury patterns. Human evaluations are error-prone, time-consuming, laborious, and subject to inter-intra-observer variability. The science and math of viewing images by a computer and human vision are entirely different.¹⁸ Automation using machine learning (ML) can be helpful in accurately assessing injury damages. ML is an area of computer engineering and artificial intelligence (AI) where computational algorithms are used in discerning meaningful patterns from imaging data.^{19,20} In automated image analysis, ML algorithms for texture detection/pattern recognition have been applied to classify cancerous lesions.^{21–23} To reduce human errors, the role of feature-based computational ML tools will be of prime importance in DILI injury assessment.

For this study, a subset of 10 common drugs associated with hepatic necrosis DILI phenotypes was selected from Lewis and Kleiner,²⁴ which provides a compilation of hepatic injury drugs and toxins. The experiment design for DILI assessment was built around feature-based (no ML) and automated ML (AutoML) models. These models were built and tested as a proof of concept for DILI injury pattern classifications. The proposed models were evaluated on drug and dose-related injury classifications by analyzing the morphometry imaging features of liver histopathology whole slide images (WSI).

MATERIALS AND METHODS

The Toxicogenomics Project–Genomics Assisted Toxicity Evaluation Systems (TG-GATES) is a publicly available open-source toxicogenomics database that stores gene expression profiles and 25 TB of digitized histopathology annotated images.²⁵ The dataset derives from *in vivo* (rat) and *in vitro* (primary rat hepatocytes, primary human hepatocytes) ex-

posure to 170 compounds. For human and rat primary hepatocytes, compounds were tested at up to four dose levels and three time-points. The datasets have been generated and analyzed over the course of 10 years of the Japanese Toxicogenomics Project, which was a joint government–private sector project organized by the National Institute of Biomedical Innovation, the National Institute of Health Sciences (NIHS), and 18 pharmaceutical companies.

The TG-GATES dataset was built up on drugs/compounds representing liver- and kidney-injuring pharmaceuticals, compounds, and chemicals. For liver imaging histopathology WSI datasets, liver sections were stained with hematoxylin and eosin and mounted on glass slides. Images of sections were converted to digital pathology images using ScanScope AT (Aperio Technologies, Inc., CA). The digital images were saved and stored in Aperio.svs format, which consists of TIFF format files with associated sample dimensions and other relevant values. Pathological information was composed of histopathological finding, topography, and grade. Annotation was performed based on a “Pathology Glossary,” a consensus-controlled vocabulary for histopathological findings for the liver and kidneys, which was originally collected by the NIHS.

For using TG-GATES in this study, which is an open-source dataset, no IRB approval was required. For this study, two models of heterogeneity assessment were built and employed: (1) a feature model, based on local imaging pixels, used to calculate fractals and lacunarity values, and (2) an AutoML model, based on deep learning, an AI algorithm. Deep learning is a subset of ML with more algorithm layers, which is further a subset of AI.

For understanding DILI features, fractal and lacunarity values were extracted from histopathology WSI. Fractal and lacunarity values are non-Euclidean, non-integer mathematical figures used to quantify structural changes, so that irregular structural and repeating patterns in an image are captured efficiently.²⁶ Figure 1 illustrates the change in fractal value D with rise in heterogeneity, which can be defined for one-dimensional, two-dimensional (2D), or three-dimensional structures. The fractal value D was estimated for a 2D irregular shape (in blue) by placing N sides (red lines) within that arbitrary shape with scaling factor r . Lacunarity is an associated property of geometric measurement, used to describe the texture and size of gaps within fractals, and represents the space filling of local structures.²⁷ In combination, fractal and lacunarity values strongly represent local structural and morphology alterations. In various cancer types, fractal and lacunarity analysis has been studied for image classification problems, including tumor and normal breast parenchyma in mammography,²⁸ lung cancer

Fig. 1. Fractal dimensions. **(a)** As heterogeneity develops, fractal values change from 2.168 to 2.864; **(b)** fractal dimension in 1D, 2D, and 3D; **(c)** fractal dimension D (space filling property) estimates the heterogeneity of curved structure by N sides (straight lines) and scaling factor r . 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional.

patterns,²⁹ human retinal vessel arborization in normal and amblyopic eyes,³⁰ bones,³¹ and malignant mesothelioma.³²

For the feature model experiment, fractal and lacunarity values of every image patch were extracted using the FracLac plugin of imagej FIJI, an open-source image processing toolbox developed by NIH,³³ illustrated in Figure 2. Eight random ROIs (regions of interests) were captured from the WSI (using the image capturing window, four ROIs per left and right lobe of liver image) per drug per dose level to equally represent and cover the injury patterns. Each capturing window was designed to have 12 grids within to further capture pixel-level information of DILI injury patterns. There is a recommended default settings of 12 grids per capturing window (which can be changed depending on the image size), which worked well on our dataset. Here, the data point represents fractal and lacunarity values corresponding to an image patch of 700×600 pixels captured at $10\times$ magnification from the WSI. The extracted fractal and lacunarity values from all these data points (image patches) were mean-averaged, respectively. Thus, each dot in a drug plot represents an average of 96 data points [8 (ROI) $\times 12$ (grids)], illustrated in Figure 3. Every patient was treated for four different dose levels, that is,

control, low, medium, and high. Thus, the image data were captured per drug per dose level. Therefore, the number of image patches, that is, N (data points per drug), is 384 [96 (data points) $\times 4$ (dose levels)]. Fractal and lacunarity values were plotted for 10 drugs, 4 dose levels (scale bar), and for 46 patients.

For the ML model, Google's AutoML Vision^{beta}, a cloud-based AutoML tool, was used (packaged with ML applications, coding, and required computer hardware). Google's AutoML Vision^{beta} enables a non-ML expert to run an ML model as per specific needs, with minimal ML expertise. Cloud-based models are supported with latest graphical processing units (GPUs), which facilitates running a custom task in a few hours (depending on the size of the dataset).

By default, for training the model, AutoML randomly divides the dataset into three separate sets: 80% training, 10% validation, and 10% out-of-box testing. The default setting can be changed anytime according to model performance. The test dataset is used for testing the model on images that the model had never seen during training (out-of-box testing). The AutoML model tries multiple parameters and algorithms during training while learning the features and

Fig. 2. Experimental workflow illustrating fractal and lacunarity values extracted with the feature model from WSI. **(a)** Eight random ROIs per WSI at 10× magnification, 12 grids per image patch capturing window. Total of $96 \times 4 = 384$ data points captured per drug. **(b)** ROIs (left) with 12 grids within (right) capturing window. ROI, region of interest; WSI, whole slide image.

injury patterns from liver image patches. To make it easy for a non-ML expert, who does not know anything about coding languages, in AutoML Vision^{beta}, model parameters are automatically adjusted and handled and are unknown to the end user. After learning the injury patterns, a validation dataset is used to validate the model. Each time a training session starts, AutoML creates a new model, and one can

select the best performing model for their custom dataset. The success and accuracy of predication depends on model design architecture, training, validation, and out-of-box testing.

The working of AutoML is based on an artificial neural network (ANN), a computational model that is based on biological brain neurons. In the ANN model, a mathematical

Fig. 3. Drug plot illustrating hepatic necrosis DILI classification using fractal and lacunarity values. Each dot is a mean-averaged value of 96 data points, that is, each dot = $[8 \text{ (ROI)} \times 12 \text{ (grids)}] = 96$ data points. Therefore, the number of data points (image patches) per drug, that is, $N = [96 \text{ (data points)} \times 4 \text{ (dose levels)}]$, is 384. Data point represents fractal and lacunarity values corresponding to an image patch of 700×600 pixels captured at 10× magnification from a whole slide image. Scale bar represents four dose levels: control, low, middle, and high. More separation in dots represents a high level of dose-dependent heterogeneity. For each drug, Pearson r , R^2 , and p -values were calculated. DILI, drug-induced liver injury.

Table 1. Drug-Induced Liver Injury at Control, Low, Medium, and High Dose Levels

Drug/compound	Image ID #	Sample ID #	Dose (mg/kg)	Time (sacrifice) (h)	Dose level	Terminal body weight (g)	Relative liver weight (%)
Acetaminophen	26824	0040041	0	24	Control	216.8	5.19
	26957	0040091LR	50	24	Low	—	—
	27039	0040142LR	300	24	Low	223.2	4.86
	27130	0040192LR	600	24	Middle	198.6	4.96
	27246	0040244LR	1,000	24	High	199.8	4.42
Bromobenzene	30151	0185041LRB	0	24	Control	201.9	4.72
	30197	0185083LRB	30	24	Low	190.6	4.39
	30231	0185121	100	24	Middle	201.1	4.62
	30270	0185161	300	24	High	195	4.86
CCl ₄	26024	0066031LRB	0	9	Control	205.5	4.53
	26098		10	24	Low	—	—
	26111	0066141	30	24	Low	198.8	4.9
	26124	0066191	100	24	Middle	196.1	4.55
	26135	0066241LRB	300	24	High	216.5	4.81
Ciprofloxacin	28048	0312041LRA	0	24	Control	189.2	5.16
	47552	0312083LRA	100	24	Low	187.9	4.01
	48250	0312125	300	24	Middle	202.5	4.43
	48320	0312161	100	24	High	185	4.74
Imipramine	16558		0	24	Control	—	—
	16635	0365081	10	24	Low	204.5	4.5
	16673	0365121LRA	30	24	Middle	207.9	4.39
	16711	0365161LRA	100	24	High	204.1	4.19
Iproniazid	18428	0318042	0	24	Control	224	4.66
	18560	0318085	6	24	Low	212.2	4.34
	18689	0318121LR	20	24	Middle	209.1	4.39
	18809	0318161LRA	60	24	High	214.8	4.32
Isoniazid	60494	0629041LRA	0	24	Control	234.8	4.76
	60506	0629081LRA	200	24	Low	225.1	4.84
	60519	0629121	600	24	Middle	212.9	4.3
	60531	0629161	2,000	24	High	219	5.33
Ketoconazole	32277	0230042LRA	0	24	Control	216	4.45
	32393	0230084LRA	10	24	Low	209.3	4.46
	32487	0230122	30	24	Middle	201.9	4.76
	32588	0230162LRA	100	24	High	211	5.36

(continued)

Table 1. Continued

Drug/compound	Image ID #	Sample ID #	Dose (mg/kg)	Time (sacrifice) (h)	Dose level	Terminal body weight (g)	Relative liver weight (%)
Methyldopa	13952	0306041	0	24	Control	206.3	4.71
	13994	0306081	60	24	Low	180.4	4.67
	14038	0306121	200	24	Middle	201.7	4.6
	14077	0306161LRA	600	24	High	164.9	3.85
Quinidine	22240	0349041LRA	0	24	Control	233.3	4.68
	22342	0349081LRA	20	24	Low	233.3	5.01
	22460	0349121LRA	60	24	Middle	215.7	4.98
	22585	0349161LRA	200	24	High	229.3	4.64
Rifampicin	4071	0094044LRB	0	24	Control	205.4	5.01
	4089	0094082LRB	20	24	Low	203.4	4.25
	4108	0094121LRB	60	24	Middle	204.6	4.23
	4128	0094161LRB	200	24	High	192.2	3.56

Species, rat; test type, *in vivo*; organ, liver; —, no information available.

figure named the artificial neuron is the basic calculation unit, which works on the principle of learning from experience and mistakes in a probabilistic and noisy environment and improving thereon. This model of ML is highly successful in industry and social network websites in accurately predicting outcomes. ML is not a new concept in cancer studies; it has been used to study various cancer types such as cancer lesion classification,³⁴ image modalities,³⁵ gastric carcinoma,³⁶ lung cancer,³⁷ classification of skin lesions,²³ cancer sequencing data,³⁸ predicting outcomes in colorectal cancer,³⁹ and drug discovery.^{40,41} ML methods have been used in primary cancer diagnosis for the identification and detection of malignant and normal image classifications.

Challenges

There are some challenges in implementing an ML model into the pathology workflow. ML is a computer engineering stream. To build an image processing model in ML, one should need a high level of expertise in coding languages such as python and MATLAB, computer vision techniques for image analysis, data processing, statistical analyses, etc. In routine clinical workflow, this is impossible for a pathologist or a digital pathology technician to learn and develop these skills, other than hire a PhD-level engineer or collaborate with an engineering group. Other challenges are

the speed of task and computer hardware. ML models are computationally very heavy to run on local computer machines. The batch job takes days to run, even for a simple image classifier, on a personal computer (PC), without any installed GPUs.

Feature engineering and data preprocessing is another barrier in designing and implementing an ML model. Feature engineering is the skill of relevant feature extraction, which includes data cleaning, background and noise filtration, artifact blocking and segmentation, etc. The model's success depends on how perfectly the task of feature engineering is performed, and what relevant features are utilized in model implementation. Irregular object boundaries and moderate tissue staining intensity are the factors that cause errors when evaluating heterogeneity by pathologists. In case of DILI, due to a rarity of this disease, grading errors depend on a pathologist's experience. Also, with a natural limitation, the human eye can't detect objects beyond a certain resolution.

To address these issues, one should need skills and trainings both in feature engineering and ML domains, which is unusual in pathology labs. To obviate these factors and to take this field out of engineering domains, AutoML is a viable alternative. Any non-ML expert, including pathologists, can design image classification models in ML to classify benign or malignant images. AutoML provides opportunities to non-ML experts to

Table 2. Data Table of Drug-Induced Liver Injury Images for Training in Automated Machine Learning

All images	1,277
Labeled	1,277
Unlabeled	0
Total labels	10
Test images	113
Average precision	0.986
Precision	92.92%
Recall	92.92%
Drug/compound	No. of images per drug
CCl ₄	128
Acetaminophen	128
Bromobenzene	128
Ciprofloxacin	128
Imipramine	128
Iproniazid	125
Isoniazid	128
Ketoconazole	128
Methyldopa	128
Quinidine	128
Precision and recall are based on a score threshold of 0.5.	

run a model without prior knowledge in any coding language. In recent years, ML has achieved considerable success and surpassed human-level performance in medical image classifications.⁴² In AutoML, it's not human, but the ANN designs and adjusts model parameters automatically.

There are a variety of AutoML models available in public domains, mostly on payment basis. Microsoft Azure AI and Google Vision AI are the two main resources equipped with all packages and libraries.

RESULTS

DILI injury classification models were build and analyzed on 10 drugs. The experiment was designed around two models to understand hepatic necrosis injury on normal rat liver due to drug intake. Drug data were built on digitally scanned WSI of patients treated for 24 h *in vivo* at four dose levels, listed in Table 1.

In the feature model, the fractal and lacunarity scores were plotted to understand heterogeneity development in image patches. More separation in dots represents a high level of dose-dependent heterogeneity, shown in Figure 3. Linear correlation and their statistical significance were analyzed between fractal and lacunarity values. The same process was repeated for all the 10 drugs. The statistical correlation was found to be high for carbon tetrachloride (CCl₄) with a Pearson correlation (*r*) of 0.9809, *p*-value of 1.6612E-06, and *R*² of 0.9582. Statistical analyses of fractal and lacunarity were analyzed. The *p*-value provides strong evidence of agreement; Pearson provides linear correlation and coefficient of determination *R*² indicates greater degree of correlation among these two dependent variables.

In AutoML, a subset of 1,277 histology images for the same 10 drugs were processed to classify DILI injury patterns. All 1,277 images were labeled, that is, provided unique identifier of the 10 drug names (mostly 128 images per label except 125 for iproniazid), as listed in Table 2. The AutoML algorithm was able to classify accurately with an average precision value of 98.6% and a score threshold of 0.5 and predict DILI injury patterns, as shown in Figure 4. AutoML provides a set of evaluation metrics for each category label and overall model performance. The area under precision/recall curve, also referred to as “average precision,” ranges normally between 0.5 (confidence threshold) and 1.0; a higher value indicates a more accurate model. The confidence threshold affects precision, recall, and true and false positive rates.

AutoML can create a confusion matrix for up to 10 labels per training dataset. A confusion matrix was created between true versus predicted labels, as shown in Figure 5; the matrix shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in orange). The ML model was able to classify true and predicted values with 100% accuracy in case of acetaminophen, bromobenzene, methyldopa, ciprofloxacin, and isoniazid. Out-of-box testing was performed on images that the model had never seen during training. The model was able to predict, with 99.8% confidence level, and correctly label unknown images; examples are shown in Supplementary Figure S1 and Figure S2.

DISCUSSION

The assessment of drug-induced injury patterns is a challenging and complex task. DILI is a rare and important liver disease, posing big concerns for the pharmaceutical industry, and is responsible for drug withdrawal from the market. Histologic manifestations of DILI can mimic chronic, acute, and all other types of liver disease patterns.

Fig. 4. Model evaluation matrix. (a) ROC precision recall curve at 0.5 threshold; (b) threshold slider to adjust precision and recall; (c) average precision, precision, and recall at 0.5 threshold. Precision and recall changes by adjusting threshold. ROC, receiver operating curve.

Confusion Matrix										
True label	Predicted label iproniazid	ketoconazole	imipramine	bromobenzene	isoniazid	CCl4	quinidine	methyldopa	acetaminophen	ciprofloxacin
iproniazid	66.7%	22.2%	-	-	-	-	11.1%	-	-	-
ketoconazole	7.7%	92.3%	-	-	-	-	-	-	-	-
imipramine	-	-	83.3%	8.3%	8.3%	-	-	-	-	-
bromobenzene	-	-	-	100.0%	-	-	-	-	-	-
isoniazid	-	-	-	-	100.0%	-	-	-	-	-
CCl4	-	-	-	-	7.1%	92.9%	-	-	-	-
quinidine	-	-	-	-	-	-	92.3%	7.7%	-	-
methyldopa	-	-	-	-	-	-	-	100.0%	-	-
acetaminophen	-	-	-	-	-	-	-	-	100.0%	-
ciprofloxacin	-	-	-	-	-	-	-	-	-	100.0%

Fig. 5. Confusion matrix: true versus predicted labels; the table shows how often the model classified each label correctly (*diagonal boxes*) and which labels were most often confused for that label.

There is no study available for DILI classification of injury patterns based on WSI. The present study is the first attempt on DILI injury classification, where a feature-based AutoML model is implemented on WSI. A study by Xu *et al.*⁴³ proposed a deep learning method for DILI pattern classifications but not on WSI. They used drug structures and compositions to understand DILI using deep learning models at molecular and chemical structural levels and bond orientations. Their proposal was on undirected graph recursive neural networks architecture for molecular structural encoding and chemical structures of single molecules.

Our study is unique and first of its kind to understand DILI injury detection and classification based on (1) histopathology WSI images and (2) the feature-based AutoML model. Also, our proposed approach is simple and novel, since even a pathologist or any non-ML expert can run this model without the knowledge of any coding language, data preprocessing, or feature engineering skills. In a routine pathology workflow, it is hard to develop the skills on writing codes, selecting and employing an ML model and computer hardware, data cleaning/preprocessing, artifact filtration, and model evaluation.

This effort is an attempt in the direction to redefine the pathology workflow in the era of ML and feature engineering. ML is rapidly growing as a process automation technology and will potentially be considered an automated diagnostic support system. Pathology labs are adopting computer-assisted diagnostic (CAD) tools, which are semi-automatic and have limited capabilities to perform feature engineering, artifact filtrations, and data preprocessing techniques. Also, CAD tools are computationally very heavy, complex to run, and incapable of selecting the most suitable model to a custom dataset.

In routine computational practice for image analysis, engineers test a couple of algorithms—from among such methods as random forest, support vector machine, *K*-mean, nearest neighbor, bag of words—to design and implement any deep learning model. To assess if the problem is eligible for ML is another challenge, that is, is the dataset eligible for clustering, regression, or a classification? It is very hard for a pathology lab to answer these questions. AutoML takes care of all these questions and make decisions itself by analyzing the custom dataset. It automatically selects a model that is already trained on a huge dataset of millions of images, which is otherwise impossible to train on a PC and needs a high-performance computing network with GPUs. The working of AutoML Vision is entirely based on transfer learning and identifying the best neural architecture for a custom dataset. Transfer learning is the process of training the model in one environment and then applying it to another unknown dataset. In the present study, a subset of

dataset was tested as a proof of concept to propose this approach as a workable model for DILI injury assessment. A successful implementation of these models for drug efficacy in the pharma industry and for clinical trials will rely on the size of DILI drug dataset, >50 drugs or so, and on all 12 phenotypes as mentioned in LiverTox.

CONCLUSIONS

In routine pathology practice, liver biopsy is a valuable diagnostic approach for DILI evaluation. Drug-associated liver injury manifests a wide range of histologic features, resembling diverse liver disease patterns, leading to liver transplant and morbidity, if not detected earlier. In pathologic diagnostics, feature engineering and ML methodologies provide unique opportunities to recognize morphologic features and so are helpful in the assessment of hepatic necrotic injury patterns. From a system perspective, a unique scoring system can be developed using fractal and lacunarity values in conjunction with ML techniques to define the etiology of DILI injury. AutoML is rapidly emerging as a tool for non-ML experts to implement a custom dataset. The present attempt is to develop a dose-dependent diagnostic support system for DILI injury classification. The use of feature-based AutoML approaches would be helpful for pharmaceutical companies, drug R&D labs, and research institutes in saving time, money, and scientific efforts during clinical trials of drug efficacy and early detection of DILI injury.

DISCLOSURE STATEMENT

No competing financial interests exist.

SUPPLEMENTARY MATERIAL

Supplementary Figure S1

Supplementary Figure S2

REFERENCES

1. Navarro VJ, Senior JR: Drug-related hepatotoxicity. *N Engl J Med* 2006;354:731–739.
2. Haque T, Sasatomi E, Hayashi PH: Drug-induced liver injury: pattern recognition and future directions. *Gut Liver* 2016;10:27–36.
3. Larson AM, Polson J, Fontana RJ, *et al.*: Acetaminophen-induced acute liver failure: results of a United States multicenter, prospective study. *Hepatology* 2005;42:1364–1372.
4. Food and Drug Administration U.S. Department of Health and Human Services: Guidance for industry drug-induced liver injury: premarketing clinical evaluation. *Drug Saf* 2009;1–25.
5. Chalasani NP, Hayashi PH, Bonkovsky HL, Navarro VJ, Lee WM, Fontana RJ: ACG clinical guideline: the diagnosis and management of idiosyncratic drug-induced liver injury. *Am J Gastroenterol* 2014;109:950–966.
6. Kleiner DE: The pathology of drug-induced liver injury. *Semin Liver Dis* 2009;29:364–372.

7. Watkins PB, Seligman PJ, Pears JS, Avigan MI, Senior JR: Using controlled clinical trials to learn more about acute drug-induced liver injury. *Hepatology* 2008;48:1680–1689.
8. Apica BS, Lee WM: Drug-induced liver injury. McManus LM, Mitchell RN (eds.), pp. 1825–1837. In: *Pathobiology of Human Disease: A Dynamic Encyclopedia of Disease Mechanisms*. Elsevier Inc., San Diego, CA, 2014.
9. Ramachandran R, Kakar S: Histological patterns in drug-induced liver disease. *J Clin Pathol* 2009;62:481–492.
10. Danan G, Benichou C: Causality assessment of adverse reactions to drugs-I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. *J Clin Epidemiol* 1993;46:1323–1330.
11. Maria VAJ, Victorino RMM: Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. *Hepatology* 1997;26:664–669.
12. Maddrey WC, Boitnott JK: Drug-induced chronic liver disease. *Gastroenterology* 1977;72:1348–1353.
13. Light DW, Lexchin JR: Pharmaceutical research and development: what do we get for all that money? *BMJ* 2012;345:e4348.
14. Kaitin KI: Deconstructing the drug development process: the new face of innovation. *Clin Pharmacol Ther* 2010;87:356–361.
15. United States National Library of Medicine: *LiverTox—Clinical and Research Information on Drug-Induced Liver Injury*. United States National Library of Medicine, Bethesda, MD, 2016.
16. Kim JW, Phongsamran PV: Drug-induced liver disease and drug use considerations in liver disease. *J Pharm Pract* 2009;22:278–289.
17. Fontana R: Acute Liver Failure Due to Drugs. *Semin Liver Dis* 2008;28:175–187.
18. Buckley D, Frisby JP, Freeman J: Lightness perception can be affected by surface curvature from stereopsis. *Perception* 1994;23:869–881.
19. Zanoni M, Arcelli Fontana F, Stella F: On applying machine learning techniques for design pattern detection. *J Syst Softw* 2015;103:102–117.
20. Urban G, Bache KM, Phan D, et al.: Deep learning for drug discovery and cancer research: automated analysis of vascularization images. *IEEE/ACM Trans Comput Biol Bioinforma* 2018 [Epub ahead of print]; DOI:10.1109/TCBB.2018.2841396.
21. Cruz JA, Wishart DS: Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2006;2:59–77.
22. Kourou K, Exarchos TP, Karamouzis MV, Fotiadis DI: Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2015;13:8–17.
23. Esteva A, Kuprel B, Novoa RA, et al.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–118.
24. Lewis JH, Kleiner DE: Hepatic injury due to drugs, herbal compounds, chemicals and toxins. In: *MacSween's Pathology of the Liver*, Burt AD, Portmann BC, Ferrell LD (eds.), pp. 645–760. Churchill Livingstone, San Diego, CA, 2012.
25. Igarashi Y, Nakatsu N, Yamashita T, et al.: Open TG-GATES: a large-scale toxicogenomics database. *Nucleic Acids Res* 2015;43(D1):D921–D927.
26. Mandelbrot B: How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science (80-)* 1967;156:636–638.
27. Plotnick RE, Gardner RH, O'Neill RV: Lacunarity indices as measures of landscape texture. *Landsc Ecol* 1993;8:201–211.
28. Guo Q, Shao J, Ruiz VF: Characterization and classification of tumor lesions using computerized fractal-based texture analysis and support vector machines in digital mammograms. *Int J Comput Assist Radiol Surg* 2009;4:11–25.
29. Lennon FE, Cianci GC, Cipriani NA, et al.: Lung cancer—a fractal viewpoint. *Nat Rev Clin Oncol* 2015;12:664–675.
30. Tălu S, Vlăduțiu C, Lupașcu CA: Characterization of human retinal vessel arborisation in normal and amblyopic eyes using multifractal analysis. *Int J Ophthalmol* 2015;8:996–1002.
31. Geraets WGM, Van der Stelt PF: Fractal properties of bone. *Dentomaxillofacial Radiol* 2000;29:144–153.
32. Lennon FE, Cianci GC, Kanteti R, et al.: Unique fractal evaluation and therapeutic implications of mitochondrial morphology in malignant mesothelioma. *Sci Rep* 2016;6:24578.
33. Karperien AL: *FracLac for Image J*. National Institute of Health, Bethesda, MD, 2013.
34. Wong D, Yip S: Machine learning classifies cancer. *Nature* 2018;555:446–447.
35. Lo C-M, (Jack) Li Y-C: Machine learning based cancer detection using various image modalities. *Comput Methods Programs Biomed* 2018;156:A1.
36. Sharma H, Zerbe N, Klempert I, Hellwich O, Hufnagel P: Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Comput Med Imaging Graph* 2017;61:2–13.
37. Coudray N, Ocampo PS, Sakellaropoulos T, et al.: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24:1559–1567.
38. Ainscough BJ, Barnell EK, Ronning P, et al.: A deep learning approach to automate refinement of somatic variant calling from cancer sequencing data. *Nat Genet* 2018;50:1735–1743.
39. Bychkov D, Linder N, Turkkilä R, et al.: Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 2018;8:3395.
40. Fleming N: How artificial intelligence is changing drug discovery. *Nature* 2018; 557:S55–S57.
41. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T: The rise of deep learning in drug discovery. *Drug Discov Today* 2018;23:1241–1250.
42. Lee JG, Jun S, Cho YW, et al.: Deep learning in medical imaging: general overview. *Korean J Radiol* 2017;18:570–584.
43. Xu Y, Dai Z, Chen F, Gao S, Pei J, Lai L: Deep learning for drug-induced liver injury. *J Chem Inf Model* 2015;55:2085–2093.

Address correspondence to:

Munish Puri, MS, PhD

Laboratory of Cancer Biology and Genetics

National Cancer Institute

National Institute of Health

9000 Rockville Pike

Bethesda, MD 20892

E-mail: twishi03@gmail.com

Abbreviations Used

AI	= artificial intelligence
ANN	= artificial neural network
AutoML	= automated machine learning
CAD	= computer-assisted diagnostic
DILI	= drug-induced liver injury
GPU	= graphical processing unit
ML	= machine learning
NIHS	= National Institute of Health Sciences
ROI	= region of interest
TG-GATES	= Toxicogenomics Project—Genomics Assisted Toxicity Evaluation System
WSI	= whole slide image