

# IBEX: An open infrastructure software platform to facilitate collaborative work in radiomics

Lifei Zhang

*Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030*

David V. Fried, Xenia J. Fave, and Luke A. Hunter

*Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas 77030*

Jinzhong Yang

*Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030*

Laurence E. Court<sup>a)</sup>

*Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas 77030*

(Received 27 October 2014; revised 15 January 2015; accepted for publication 2 February 2015; published 25 February 2015)

**Purpose:** Radiomics, which is the high-throughput extraction and analysis of quantitative image features, has been shown to have considerable potential to quantify the tumor phenotype. However, at present, a lack of software infrastructure has impeded the development of radiomics and its applications. Therefore, the authors developed the imaging biomarker explorer (IBEX), an open infrastructure software platform that flexibly supports common radiomics workflow tasks such as multimodality image data import and review, development of feature extraction algorithms, model validation, and consistent data sharing among multiple institutions.

**Methods:** The IBEX software package was developed using the MATLAB and C/C++ programming languages. The software architecture deploys the modern model-view-controller, unit testing, and function handle programming concepts to isolate each quantitative imaging analysis task, to validate if their relevant data and algorithms are fit for use, and to plug in new modules. On one hand, IBEX is self-contained and ready to use: it has implemented common data importers, common image filters, and common feature extraction algorithms. On the other hand, IBEX provides an integrated development environment on top of MATLAB and C/C++, so users are not limited to its built-in functions. In the IBEX developer studio, users can plug in, debug, and test new algorithms, extending IBEX's functionality. IBEX also supports quality assurance for data and feature algorithms: image data, regions of interest, and feature algorithm-related data can be reviewed, validated, and/or modified. More importantly, two key elements in collaborative workflows, the consistency of data sharing and the reproducibility of calculation result, are embedded in the IBEX workflow: image data, feature algorithms, and model validation including newly developed ones from different users can be easily and consistently shared so that results can be more easily reproduced between institutions.

**Results:** Researchers with a variety of technical skill levels, including radiation oncologists, physicists, and computer scientists, have found the IBEX software to be intuitive, powerful, and easy to use. IBEX can be run at any computer with the windows operating system and 1GB RAM. The authors fully validated the implementation of all importers, preprocessing algorithms, and feature extraction algorithms. Windows version 1.0 beta of stand-alone IBEX and IBEX's source code can be downloaded.

**Conclusions:** The authors successfully implemented IBEX, an open infrastructure software platform that streamlines common radiomics workflow tasks. Its transparency, flexibility, and portability can greatly accelerate the pace of radiomics research and pave the way toward successful clinical translation. © 2015 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4908210>]

Key words: radiomics, quantitative imaging analysis, infrastructure software, collaborative work

## 1. INTRODUCTION

Patients receive an ever increasing number of multimodality imaging procedures, such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). The use and role of medical images has greatly expanded from primarily as a diagnostic tool to include

a more central role in the context of individualized medicine.<sup>1-5</sup> At present, the effective utilization of this large amount of medical imaging data is still challenging. Recently, there is an increased interest in the use of quantitative imaging methods to both improve tumor diagnosis and act as proxies of genetics and tumor response. With these improvements, the overall goal is to better inform and enhance clinical decision

making.<sup>6–19</sup> One important advancement in quantitative imaging analysis is the concept of “radiomics.” Radiomics is the high-throughput extraction and analysis of quantitative imaging features from medical images.<sup>20,21</sup> Previous work has shown that radiomics can be used to create improved prediction algorithms for various clinically relevant metrics and endpoints.<sup>22–25</sup>

The lack of an open infrastructure software platform, however, has made previous radiomics research difficult to share and validate between institutions. Image features with the same name may be implemented differently by different groups. For example, the number of bins used for calculating histograms may vary, as may the use of image interpolation. These differences mean that independent validation of published work is difficult. As a result, the translation of radiomics research findings into improved clinical practices has been notably impeded. There is, therefore, a need for an open infrastructure software platform that is available for all researchers. Currently, no infrastructure software platforms are available to flexibly support common quantitative imaging analysis tasks such as multimodality image data import and review, development and calculation of feature extraction algorithms, model validation, and consistent data sharing among multiple institutions to assess reproducibility. A Computational Environment for Radiotherapy Research (CERR) publication<sup>26</sup> states that reproducibility is a *key element* of the scientific method. This has been difficult to achieve with previous radiomics implementations. Some publicly available software programs do however exist for specific image feature analysis. For example, Chang-Gung Image Texture Analysis (CGTIA)<sup>27</sup> is an open-source software package for quantifying tumor heterogeneity with PET images. Also, originally designed for MRI texture analysis, MaZda (Ref. 28) is another software package with SDK support. Because of their intended use, both software packages are limited in their functionality or scope. For example, in neither is there a simple way for the user to implement new types of image features. Modifying feature extraction parameters, reviewing, validating intermediate data and results are also challenging using these two software packages. Straightforward multi-institutional reproduction of results is not included in their workflows. Within the field of radiation oncology, CERR demonstrates a successful example of open-source software used for collaborative work, and can be used as a template for the development of similar software geared toward radiomics.

We developed the imaging biomarker explorer (IBEX) software package as an open infrastructure software platform to flexibly support common radiomics workflow tasks such as multimodality image data import and review, development of feature extraction algorithms, model validation, and consistent data sharing among multiple institutions. IBEX is used for research only. On one hand, IBEX is a self-contained and ready-to-use radiomics software, with preimplemented typical data importers, image filters, and feature extraction algorithms. On the other hand, the advanced research developers can extend IBEX’s functionality. IBEX provides an integrated development environment on top of the MATLAB (MathWorks, Natick, MA) and C/C++ programming languages. Users are not limited to

IBEX’s built-in functions: in the IBEX developer studio, users can plug in, debug, and test new algorithms, extending the program’s functionality. IBEX also supports quality assurance for data and feature extraction algorithms: image data, ROIs, and feature algorithm-related data can be reviewed, validated, and modified. Critically, image data, feature extraction algorithms, and model validation can be anonymized and be easily and consistently shared so that users from different institutions can reproduce the results of radiomics workflows. Finally, windows version 1.0 beta of stand-alone IBEX without the requirement of MATLAB license can be freely downloaded at [http://bit.ly/IBEX\\_MDAnderson](http://bit.ly/IBEX_MDAnderson). The source-code version of IBEX can be downloaded at [http://bit.ly/IBEXSrc\\_MDAnderson](http://bit.ly/IBEXSrc_MDAnderson) for free. Both versions of IBEX can be shipped in compact disc form as well.

The alpha version of IBEX was developed by Hunter *et al.*<sup>14</sup> for in-house radiomics analysis. The current 1.0 beta version of IBEX discussed in this paper was created from scratch in order to increase performance, improve ease of use, and extend functionality. Most importantly, compared to the prior version, the current IBEX version has been engineered to have greatly increased modularity and robustness, allowing for it to be used collaboratively across multiple institutions.

## 2. DESCRIPTION OF IBEX

### 2.A. Software architecture

IBEX is written using MATLAB 2011a, 32-bit programming environment. To overcome poor memory management for large matrices, many three-dimensional (3D) image analysis modules are written in C/C++ and called by MATLAB via the MATLAB Executable (MEX) interface. IBEX consists of a suite of component-based application and development tools for applying, sharing, and building reliable and reproducible quantitative image analysis algorithms.

To achieve the goal of being an open infrastructure software platform, three modern programming concepts—model-view-controller (MVC),<sup>29</sup> unit testing,<sup>30</sup> and function handles—were deployed as shown in Fig. 1.

The MVC concept is implemented to isolate each task. By implementing a unit testing concept, users are able to validate if their relevant data and algorithms are fit for use. Function handles are widely employed in the supplied developer studio of IBEX where users can easily plug in their own algorithms into IBEX. The MVC *View Component* represents the workspace of reviewing multimodality images with delineated structures (if available). The MVC *Controller Component* represents image preprocessing and feature extraction algorithms. The MVC *Model Component* represents the predictive model formula and parameters. Because of the unit testing implementation, users have the option of reviewing the corresponding result at each stage to check on the quality of the data and algorithms. Although IBEX is self-contained and has standard algorithms and modules for a typical radiomics workflow, it is an open system, so additional algorithms and models can be easily added by defining them in library files in the IBEX developer studio. Thanks to the MVC technique, a complete

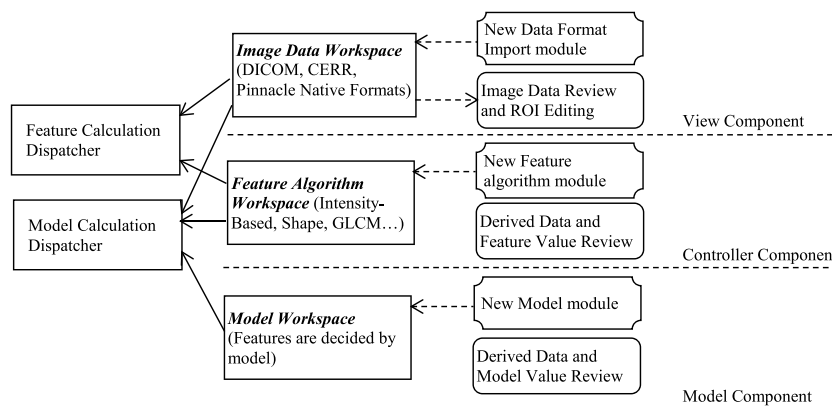


FIG. 1. The IBEX architecture. MVC, unit testing, and function handle programming concepts are deployed to isolate each task, test algorithms, plug in new algorithms, share data, and reproduce data easily and consistently.

model can be exported or imported easily, including necessary data such as preprocessing algorithms, feature extraction algorithms, model formulas, and model parameters. This greatly helps maintain data consistency and result reproducibility when outside institutions attempt to validate feature extraction algorithms and response models.

The IBEX workflow is shown in Fig. 2. The *IBEX Database* is a local store of patient images with associated data and ROIs. Regular users begin their workflow (Step #1) by importing patient data into the *IBEX Database* using Digital Imaging and Communications in Medicine (DICOM)<sup>31</sup> format data importer or PINNACLE (Philips Radiation Oncology Systems, Fitchburg, WI) native format data importer. The *Data Set* is a local store of images that are subportions of images previously added to the *IBEX Database*. To create subimages to populate

the *Data Set* (Step #2), users open an image from the *IBEX Database*; review the image and its associated ROIs; modify or create ROIs if desired; and specify which ROIs to apply to the image to obtain a subimage (multiple subimages can be generated from the same patient image by applying different ROIs). The *Feature Set* is a local store of the features that the user wishes to have extracted from a subimage. IBEX organizes a variety of features into several feature categories based on feature’s nature. For example, all intensity histogram related features belong to the feature category “IntensityHistogram.” Feature category code computes the parent data (the parent data correspond to the histogram data for the feature category IntensityHistogram) and sends the parent data to feature extraction algorithm code to compute the value of each individual feature (features correspond to kurtosis, skewness, etc.

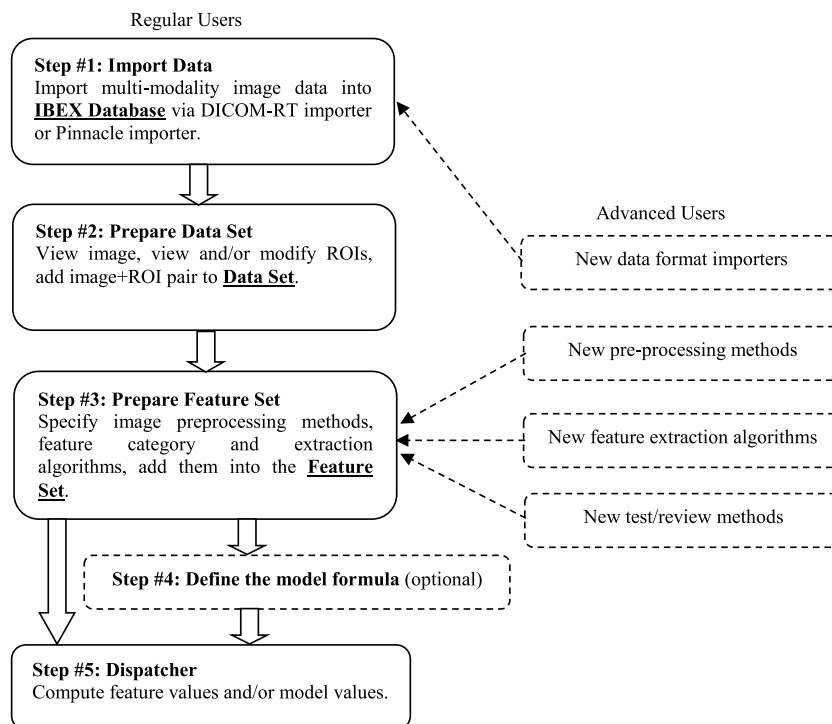


FIG. 2. The IBEX workflow. Regular users import data, prepare the data set and feature set, specify the model formula, and compute the feature value and/or model value. Advanced users can plug in new data format importers, preprocessing methods, feature algorithms, and test review methods using the IBEX developer studio.

for feature category IntensityHistogram). Users add features to the *Feature Set* (Step #3) by specifying image preprocessing algorithm(s), feature category, and feature extraction algorithm(s). Algorithm results can be reviewed via testing (optional). Users can then specify a model formula (Step #4) if desired. To complete the workflow (Step #5), users specify the *Data Set* and the *Feature Set* created in the previous steps and direct IBEX to compute the feature values and/or model values. The steps above describe how to use IBEX's built-in functions. Advanced IBEX users can use the IBEX developer studio to plug in new data format importers, preprocessing methods, feature extraction algorithms, and test/review methods.

## 2.B. Image data workspace

The main purpose of the *Image Data Workspace* in IBEX is to create subimages (image/ROI pairs) to add to a *Data Set*. Each item within a *Data Set* contains the basic information about an image and ROI pair, such as the imaging modality, medical record number (MRN), ROI statistics, voxel and image information, and item creation time. The *Data Set* also stores ROI contours, ROI binary masks, and image data in the ROI bounding box.

To prepare each *Data Set*, IBEX includes the functionalities of importing patient data, reviewing images and ROIs, modifying or creating ROIs if necessary, appending *Data Set* items by adding image and ROI pairs. In compliance with the unit testing philosophy, IBEX also supports reviewing and modifying *Data Set* items in the current workspace.

The current version of IBEX provides DICOM data and PINNACLE native data importers. PINNACLE native data are the raw data used by the PINNACLE treatment planning system (TPS). DICOM data may originate from numerous sources, including the majority of radiotherapy treatment planning systems, such as Eclipse (Varian Medical Systems, Palo Alto, CA) and PINNACLE, many free image viewers/editors, such as 3D SLICER (<http://www.slicer.org>), and the majority of commercial segmentation systems, such as MIMvista (MIM Software, Inc., Cleveland, OH), Velocity (Varian Medical Systems, Palo Alto, CA), and Mirada (Mirada Medical, Oxford, UK). If a computer running IBEX has access to a PINNACLE postgres database and data storage, IBEX can be configured to retrieve data in the PINNACLE native format directly from storage. For a DICOM importer, it first reads all the files in a configured DICOM input directory, then sorts and organizes the DICOM data according to the unique identifier (UID), and then lists all the patients available for import. As part of the unit testing implementation, the Details list box in the DICOM data importer describes the patient information and any related plan, ROI, and image information. Figure 3 shows an example of a DICOM data importer. When importing a patient's DICOM data, IBEX converts the data into the PINNACLE native format. If DICOM imaging data were obtained using PET, IBEX automatically computes the standardized uptake value from the DICOM PET raw uptake value if all the necessary radiopharmaceutical dose information is available. In addition to conversion, IBEX also dumps all DICOM file information into the DICOMInfo folder to retain all the information from DICOM files.

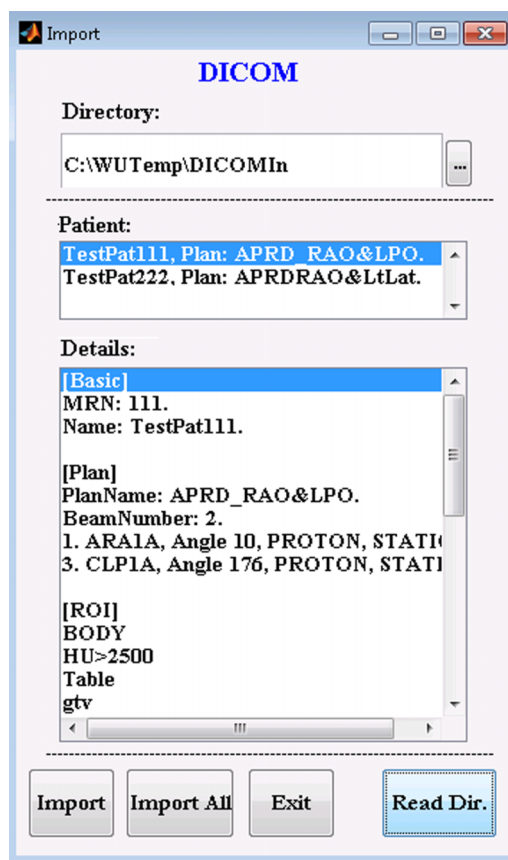


Fig. 3. Example of a DICOM data importer. The importer sorts and organizes DICOM data based on the relationship among MRNs, instance UIDs, study UIDs, series UIDs, and frame UIDs, and then lists all the available patients that could be imported. The Details list box describes the detailed patient information for verification.

IBEX cannot connect to any PACS and RIS/HIS at present—the images will first need to be exported from the PACs and then imported into IBEX. The current build-in IBEX importers cannot import data from non-DICOM or non-PINNACLE objects. However, users can plug-in their own customized data importers through IBEX *Developer Studio* to import those non-DICOM or non-PINNACLE data.

In the IBEX *Image Data Workspace*, users insert *Data Set* items by specifying image and ROI pairs. Figure 4 is a screenshot of the IBEX *Image Data Workspace*. This workspace supplies the multimodality image viewer for axial, coronal, and sagittal orientations and the ROI editor. Users can navigate to the different image slices, zoom images in and out, quickly go to the corresponding anatomy using the intersection tool, measure the distance, check image intensity values, manually set window/level, select the preset window and/or level setting, and select the preset color map. As part of unit testing implementation, ROIs can be overlaid on images in three orientations to verify contours. If a ROI must be modified, the user can employ the ROI editor to create a new ROI, copy the existing ROI, delete the ROI, nudge contours, delete contours, draw contours by clicking points, freely draw contours, or interpolate contours. Figure 5 is a screenshot of the ROI editor tools in IBEX.

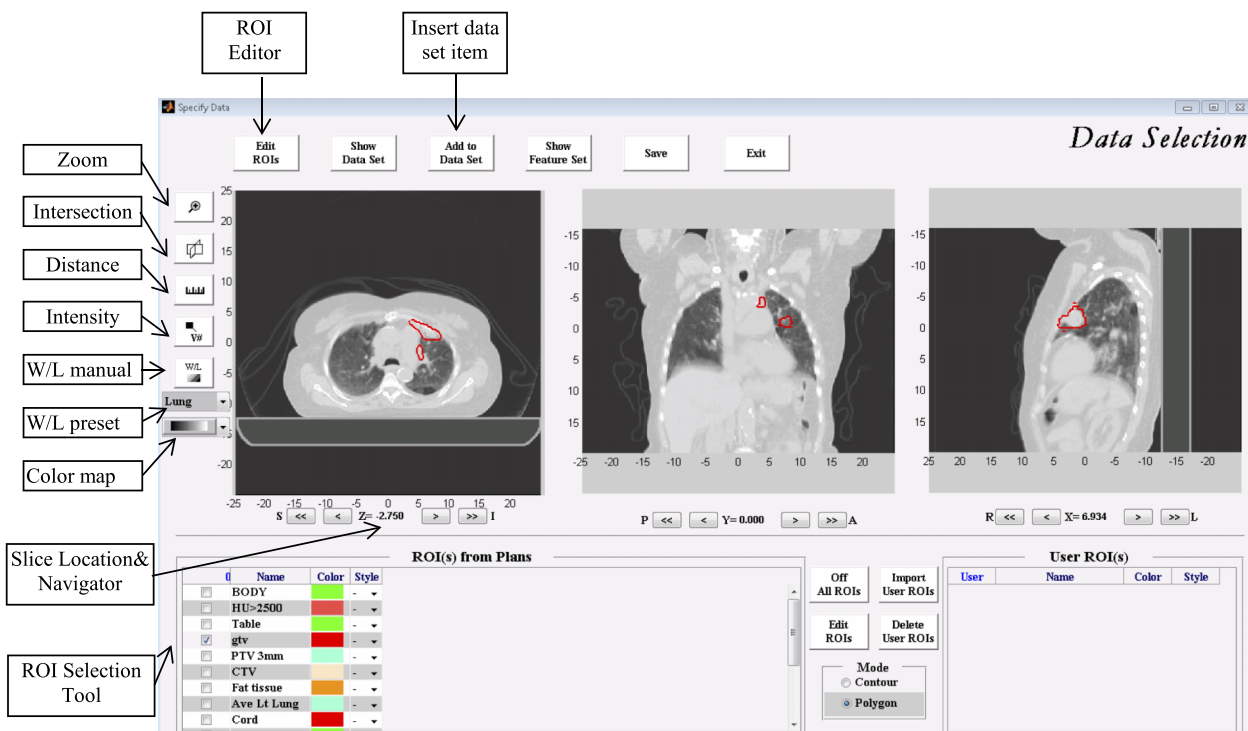


FIG. 4. The iBEX image data workspace. The main purpose of this workspace is to insert data set items by specifying image and ROI pairs. Image data can be viewed in axial, coronal, and sagittal orientations. ROIs can be overlaid on images and modified if necessary. Users can navigate to different image slices, zoom images in and out, quickly view the corresponding anatomy using the intersection tool, measure the distance, check the image intensity value, manually set window/level, select the preset window/level setting, and select the preset color map.

### 2.C. Feature algorithm workspace

The main purpose of the *Feature Algorithm Workspace* in iBEX is to prepare the *Feature Set* by specifying image preprocessing algorithms, feature category, and feature extraction algorithms. Each *Feature Set* item contains preprocessing methods and their parameters, the feature category and its parameters, feature extraction algorithms and their parameters, and the current feature set information (such as comments and its creation date). Figure 6 is a screenshot of the iBEX *Feature Algorithm Workspace*.

In the *Feature Algorithm Workspace*, users first specify the image preprocessing algorithms applied to the image. Users can apply multiple preprocessing algorithms in any order. Multiple preprocessing algorithms work in a pipeline style. Table I lists all of the preprocessing algorithms currently available in the current version of iBEX. Users then specify the feature category and its feature extraction algorithms. The feature categories and feature extraction algorithms currently available in iBEX are listed in Table II. As part of unit testing imple-

mentation, users can review and modify algorithm parameters using a parameter modification graphical user interface (GUI). Furthermore, by clicking the test button in the workspace, users can test the algorithm and review the intermediate data and feature calculation result. Figure 7 shows an example of testing the feature “Kurtosis” in the category IntensityHistogram. In the review window, users can check the original and preprocessed images, feature values, and contours.

The *Feature Algorithm Workspace* is also self-documented. Each feature name is self-explanatory, indicating what feature it is. For example, the feature “ConvexHullVolume3D” in the category “Shape” means that the volume of the ROI convex hull is calculated according to the 3D connectivity of adjacent voxels in the binary masks. A detailed description of the algorithm and its parameters is easily accessed by clicking the help button on the parameter modification GUI as shown in Fig. 8.

The feature categories “GrayLevelCooccurrenceMatrix”<sup>32,33</sup> and “NeighborhoodIntensityDifference”<sup>34</sup> in the *Feature Algorithm Workspace* are implemented in both two and a half

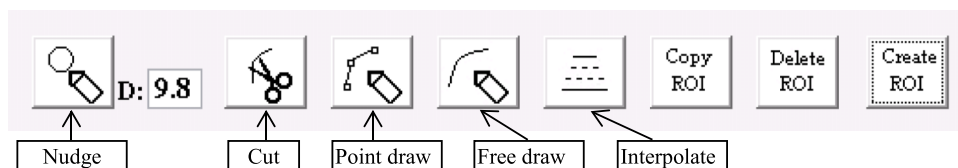


FIG. 5. The ROI editor tools in iBEX. Users can use the ROI editor to create new ROIs, copy existing ROIs, delete ROIs, nudge contours, delete contours, draw contours by clicking points, freely draw contours, and interpolate contours.

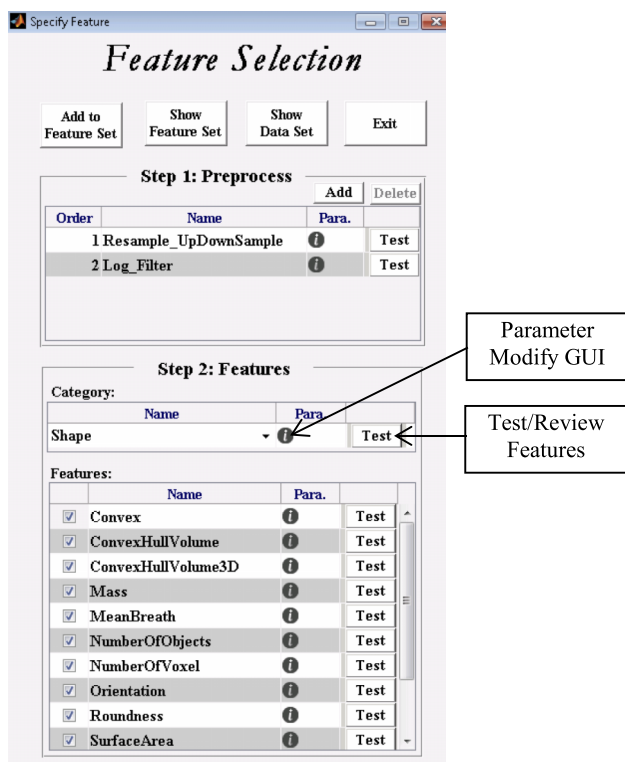


FIG. 6. The feature algorithm workspace in IBEX. The main purpose of this workspace is to prepare the feature set by specifying the image preprocessing algorithms, feature category, and feature algorithms.

dimension (2.5D) and 3D versions. This is done in consideration of the fact that most image data are in finer resolution in one orientation than in others. For example, the

feature “GrayLevelCooccurrenceMatrix25” computes the co-occurrence of individual intensity pairs in 2D directions in slice by slice manner. Next, the gray-level co-occurrence matrix is the summation of the co-occurrence of individual intensity pairs in all 2D image slices. In contrast, the feature “GrayLevelCooccurrenceMatrix3” directly computes the GLCM as the co-occurrence of individual intensity pairs in the 3D directions. Similarly, in the feature “NeighborhoodIntensityDifference25,” the NID matrix is computed with the voxel’s neighborhood defined in 2D, whereas in the feature “NeighborhoodIntensityDifference3,” the neighborhood is defined in 3D. In the feature “GrayLevelRunLengthMatrix25,”<sup>35,36</sup> the RLM is computed in 2D.

It is important to set the appropriate algorithm parameters for different modality images. The default parameters are set to be suitable for CT modality images. Figure 9 shows histograms created using the different parameters for one PET image set [Fig. 9(A)]. If the CT parameters in Fig. 9(B) are *inappropriately* applied to the PET data, then the histogram is erroneously compressed into one bin location [Fig. 9(D)]. However, by correctly selected PET-appropriate parameters, the appropriate histogram results are generated [Fig. 9(E)]. Note, consistent with the PINNACLE treatment planning system, IBEX uses a CT number where water is given a value of 1000.

### 2.D. Model workspace

The main purpose of the *Model Workspace* in IBEX is to prepare the model formula by specifying the expression, features, and parameters of each item. In the current version of IBEX, the model formula (i.e., the formula that adds different features with different weights to give an outcome prediction) is simply

TABLE I. The image preprocessing algorithms available in IBEX.

Purpose	Preprocessing name	Comment	References
Image smoothing	Average_Smooth		
	EdgePreserve_Smooth3D		
	Gaussian_Smooth		11–14 and 16
	Gaussian_Smooth3D		
	Median_Smooth		
	Wiener_Smooth		11
Image enhancement	AdaptHistEqualization_Enhance3D		
	HistEqualization_Enhance		
	Sharp_Enhance		
Image deblur	Blind_Deblur		
	Gaussian_Deblur		
Change enhancement	Laplacian_Filter		11–14 and 16
	Log_Filter		11–14 and 16
	XEdge_Enhance		
	YEdge_Enhance		
Resample	Resample_UpDownSample		9
	Resample_VoxelSize		9
Miscellaneous	Threshold_Image_Mask		11 and 15
	Threshold_Mask		11 and 15
	BitDepthRescale_Range	Change dynamic range	11 and 15

TABLE II. The feature extraction algorithms available in IBEX.

Category	Feature name	Comment	References
Shape	Compactness1		33
	Compactness2		33
	Max3DDiameter		33
	SphericalDisproportion		33
	Sphericity		33
	Volume		7 and 15
	SurfaceArea		7
	SurfaceAreaDensity		
	Mass	Useful for CT only	7
	Convex		7
	ConvexHullVolume		
	ConvexHullVolume3D		
	MeanBreadth		
	Orientation		7
	Roundness		7
	NumberOfObjects		
NumberOfVoxel		7	
IntensityDirect	Energy		33
	RootMeanSquare		33
	Variance		33
	Kurtosis		7, 9, 11, and 15
	Skewness		7, 9, 11, and 15
	Range		9
	Percentile		9 and 15
	Quantile		9
	InterQuartileRange		9
	GlobalEntropy		7, 9, 11, 12, and 15,
	GlobalUniformity		11–14 and 16
	GlobalMax		9 and 15
	GlobalMin		9 and 15
	GlobalMean		7, 9, 11–13, and 15
	GlobalMedian		9 and 15
	GlobalStd		7, 9, 11–13, and 15,
	MeanAbsoluteDeviation		9
	MedianAbsoluteDeviation		
	LocalEntropy/Range/StdMax		
LocalEntropy/Range/StdMin			
LocalEntropy/Range/StdMean			
LocalEntropy/Range/StdMedian			
LocalEntropy/Range/StdStd			
IntensityHistogram	Kurtosis		7, 9, 11, and 15
	Skewness		7, 9, 11, and 15
	Range		
	Percentile		9 and 15
	PercentileArea		
	Quantile		9
	InterQuartileRange		9
	AutoCorrelation		32 and 33
	ClusterProminence		32 and 33
	ClusterShade		32 and 33
	CluserTendency		32 and 33
	DifferenceEntropy		32 and 33
	Dissimilarity		32 and 33
	Entropy		32 and 33
Homogeneity2		32 and 33	
InformationMeasureCorr1		32 and 33	

TABLE II. (Continued).

Category	Feature name	Comment	References
GrayLevelCooccurrenceMatrix25 GrayLevelCooccurrenceMatrix3	InformationMeasureCorr2	25:=GLCM is computed	32 and 33
	InverseDiffMomentNorm	from all 2D image slices	32 and 33
	InverseDiffNorm	3:=GLCM is computed	32 and 33
	InverseVariance	from 3D image matrix	32 and 33
	MaxProbability		32 and 33
	SumAverage		32 and 33
	SumEntropy		32 and 33
	SumVariance		32 and 33
	Variance		32 and 33
	Contrast		7, 9, 11, 15, 28, 32, and 33
	Correlation		7, 9, 11, 15, 28, 32, and 33
	Energy		7, 9, 11, 15, 28, 32, and 33
	Homogeneity		7, 9, 11, 15, 28, 32, and 33
NeighborIntensityDifference25 NeighborIntensityDifference3	Busyness	25:= neighborhood	11, 23, 29, and 34
	Coarseness	intensity difference (NID)	11, 23, 29, and 34
	Complexity	is computed from all 2D	23, 29, and 34
	Contrast	image slices	11, 23, 29, and 34
	TextureStrength	3:=NID is computed from 3D image matrix	23, 29, and 34
GrayLevelRunLengthMatrix25	GrayLevelNonuniformity		7, 15, 30, 31, 35, and 36
	HighGrayLevelRunEmpha		7, 30, 31, 35, and 36
	LongRunEmphasis		7, 15, 30, 31, 35, and 36
	LongRunHighGrayLevelEmpha	25:=run-length matrix	7, 30, 31, 35, and 36
	LongRunLowGrayLevelEmpha	(RLM) is computed from	7, 30, 31, 35, and 36
	LowGrayLevelRunEmpha	all 2D image slices	7, 30, 31, 35, and 36
	RunLengthNonuniformity	25:=RLM is computed	7, 15, 30, 31, 35, and 36
	RunPercentage	from all 2D image slices	7, 15, 30, 31, 35, and 36
	ShortRunEmphasis		7, 15, 30, 31, 35, and 36
	ShortRunHighGrayLevelEmpha		7, 30, 31, 35, and 36
ShortRunLowGrayLevelEmpha		7, 30, 31, 35, and 36	
IntensityHistogramGaussFit	GaussAmplitude		
	GaussArea		
	GaussMean		
	GaussStd		
	NumberOfGauss		

defined in an ASCII text file to make it readable and easily shared. Defining the model formula naturally indicates which features are used in the model. Modeling is the informatics analysis of features. Models can be generated for different applications, such as tumor diagnosis, tumor staging, gene prediction, and outcome prediction. Developing a good model and selecting appropriate model features are beyond the scope of this report. Examples of model development have been described by several authors.<sup>7,11</sup> That is, the predictive models must be developed outside of *IBEX*. The image features (including all necessary parameters) and model coefficients can then be added to *IBEX* for the purpose of internal or independent validation, etc.

## 2.E. Computation dispatcher

The Computation dispatcher (see Fig. 3, Step #5) in *IBEX* is used to compute the feature or model values. Users first

specify the data set, feature set, and/or model. The dispatcher engine then computes the feature and/or model value. Last, the dispatcher writes the result along with the information of data set, feature set, and model into one Excel spreadsheet (Microsoft Corporation, Redmond, WA). Users can then use this common format to import the result into the statistical program (SPSS, R, SAS, etc.) that they prefer. To comply with unit testing implementation, the information in the Excel file contains the data set item description, features' names and parameters, and model formula so that users can reproduce the results and determine what was used to generate the results. Thanks to MVC concept implementation, the data sets, feature sets, and models are relatively independent of one another. Thus, one data set can be applied to different feature sets and models, one feature set can be applied to different data sets and models, and one model can be applied to different feature and data sets. This independence in the *IBEX* workspaces enables *IBEX* to serve as the infrastructure platform for testing



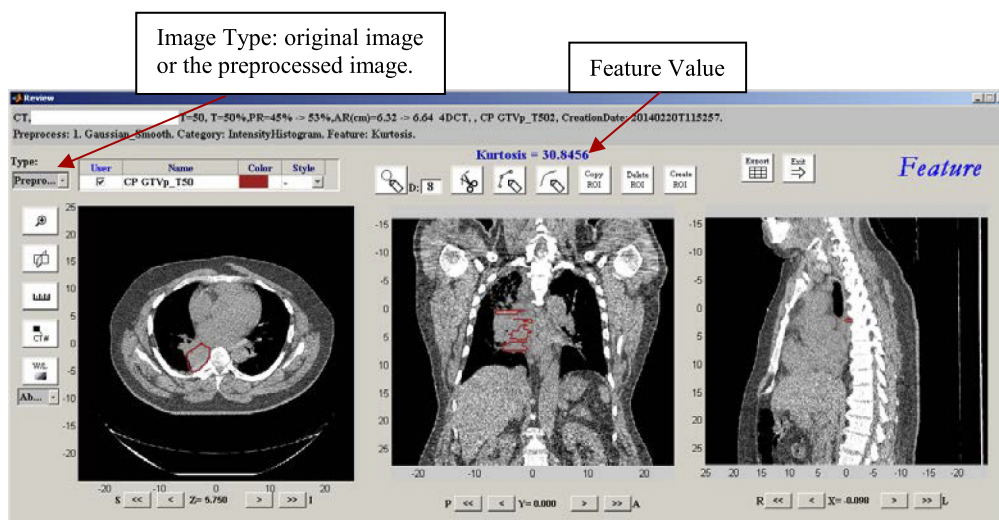


FIG. 7. A testing GUI in IBEX. At each stage (import, preprocessing, and feature calculation), users have the option of reviewing the corresponding results and intermediate data.

and developing feature algorithms and models for quantitative imaging analysis.

idea of plug in input arguments, and then modify and enrich the skeleton code to meet their purposes.

### 2.F. Developer studio/extensibility

The IBEX developer studio enables users to extend the functionality of the software. In the developer studio, advanced users can plug in new data importers for any data format, new preprocessing algorithms, new feature algorithms, and new test/review functions. The IBEX plug-in feature is based heavily on the MATLAB function handle technique. The IBEX developer studio works in the same way as Visual Studio (Microsoft Corporation). Depending on the type of plug-in, the developer studio generates the skeleton code with simple functions and puts this code under the designated directory for the IBEX platform to recognize. The skeleton code itself is ready to use. Advanced users can first run the skeleton code to get an

### 2.G. Reproducibility

Because of the MVC architecture, interinstitutional comparison and reproducibility can be easily provided by IBEX. Data and feature sets are stored in the individual MATLAB MAT files, and models are stored in readable ASCII files. Data sets, feature sets, and models are all self-contained, including all information necessary for the IBEX computation dispatcher to calculate the result of feature and/or model. The IBEX users anonymize and export their own data set, feature set, and/or model files. An IBEX user from a different institution can then import these files into the IBEX database and then compute the result of feature and/or model. The result can be reproduced, as all the data are shared consistently among institutions.

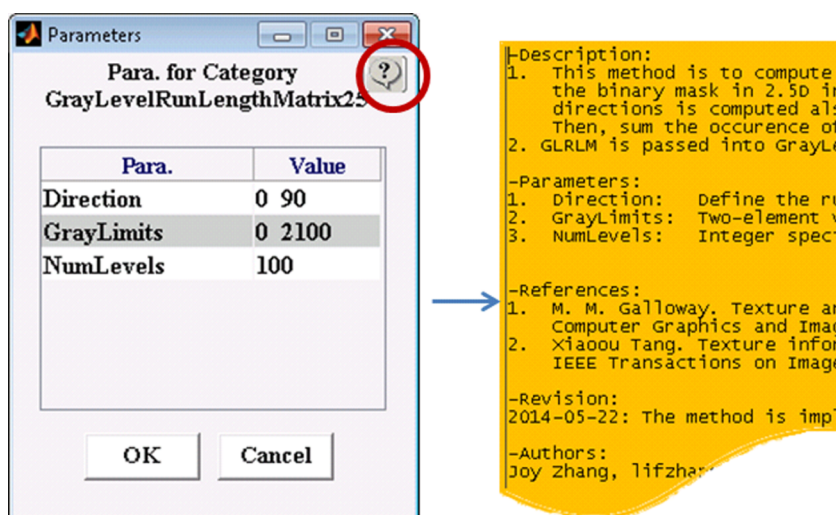


FIG. 8. Self-documented algorithm in IBEX. The algorithm and feature name are self-explained. The description of the algorithm and its parameters can be easily accessed using the help button on the parameter modification GUI (circled in red).

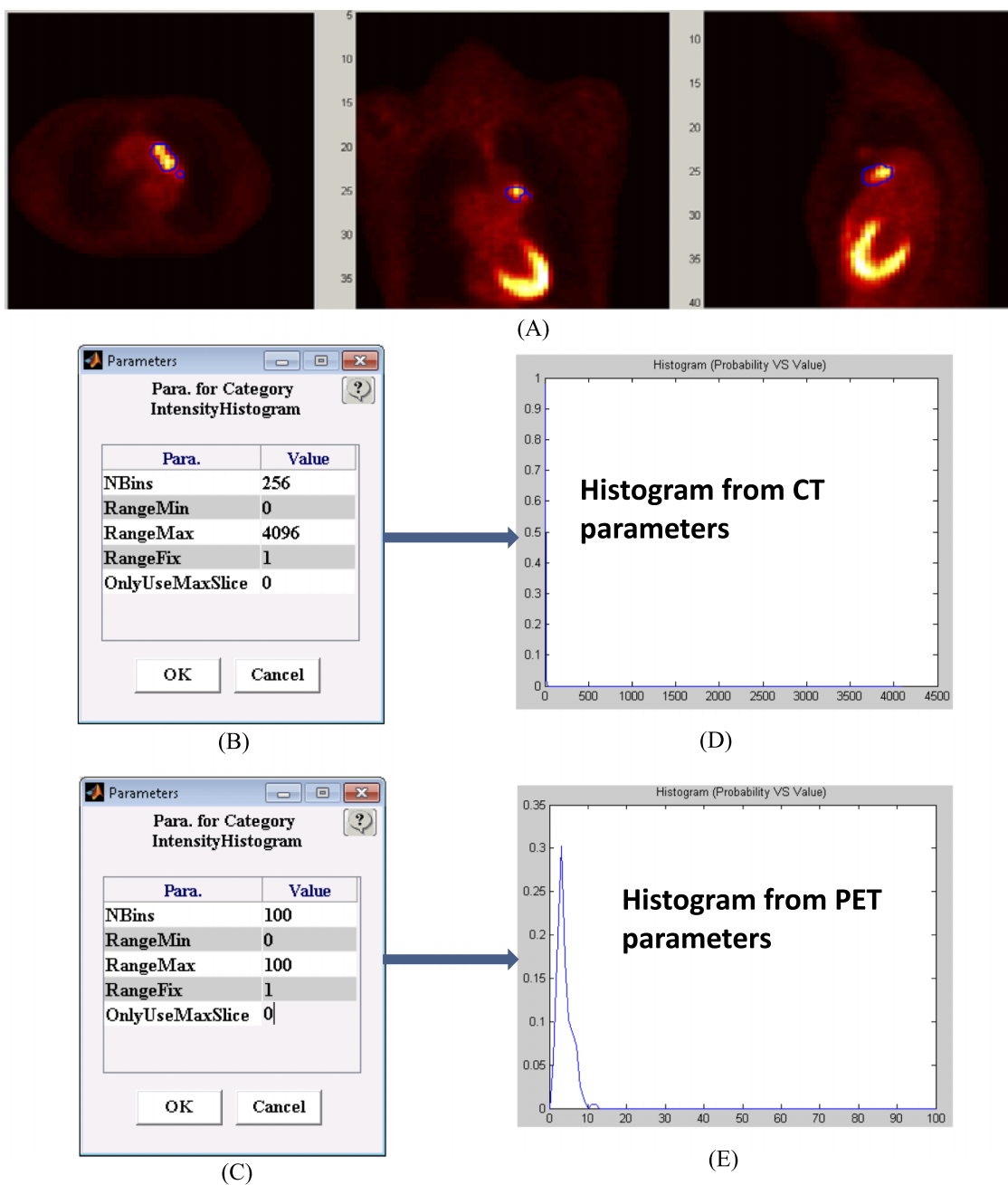


Fig. 9. The appropriate algorithm parameters for different modality images. (A) PET image. (B) CT-type parameters. (C) PET-type parameters. (D) Histogram from CT-type parameters that is meaningless and squeezed into one bin. (E) Histogram from PET-type parameters. The PET-type parameters zoom in on a CT-type histogram and can provide meaningful results for a PET image.

The second user can double-check the first user’s algorithms by examining the parameters and reviewing the intermediate data. Data anonymization can be done in several scenarios: Users have the option to anonymize their data when data are imported into IBEX; users can anonymize patients in the IBEX database; data sets created in IBEX can be anonymized; in the data workspace, the user has a tool to anonymize the ROI data.

### 2.H. Quality assurance/reliability

Thanks to the implementation of unit testing philosophy, IBEX users can review the relevant data at each stage involved

in the feature calculation. Specifically, IBEX provides the GUIs for users to review image data, review and modify ROIs and algorithm parameters, read algorithm descriptions, test algorithms, review intermediate and final results for algorithms, and review model formulas. IBEX itself supports reviewing 3D and 2D matrices, single values, gray-level co-occurrence matrices, curves, meshes, and layers along with the image display. Furthermore, users can even plug in their own review callback functions to customize the review requirements. All of these capabilities enable users to perform quality assurance for their image data, IBEX’s built-in algorithms, users’ plug-ins, and models.

## 2.I. Testing

The implementation of data importers, preprocessing algorithms, and feature extraction algorithms in *IBEX* was validated using commercial and free software. Specifically, the *IBEX* DICOM importer was compared with DICOM importers in the *ECLIPSE* and *PINNACLE* TPSs. Also, the *IBEX* *PINNACLE* importer was compared with the *PINNACLE* TPS database. *IBEX* preprocessing algorithms were validated against *MATLAB*'s built-in functions and *CERR* implementation qualitatively by visually reviewing the preprocessed images. Software developers and physics users visually reviewed the preprocessed images for each modality (5+ images for CT, MRI, and PET modalities). This qualitative comparison was subjective, with the users visually searching the differences in the preprocessed images created by *IBEX*, *MATLAB*, and *CERR*. For the purpose of the quantitative validation, we created four digital sphere phantoms with one known volume size (volume = 65.3 cm<sup>3</sup>), one known mean intensity value (mean = 1025), and four different intensity standard deviations (SD = 25, 47, 50, and 75). By comparing with the known value, the average volume differences were 0.15, 0.03, 0.69, and 1.28 cm<sup>3</sup> from *IBEX*, *PINNACLE* TPS, *ECLIPSE* TPS, and *CGITA*; the average intensity mean difference was 0.20, 0.16, and 0.21 from *IBEX*, *PINNACLE* TPS, and *CGITA*, respectively; the average intensity standard deviation difference was 0.23, 0.22, and 0.24 from *IBEX*, *PINNACLE* TPS, and *CGITA*, respectively. Feature values of Kurtosis and skewness from *IBEX* on these four digital phantoms were compared with those from *CGITA*. The average Kurtosis difference and skewness difference are 0.02 and 0.00. We qualitatively validated feature algorithm implementation for categories GLCM, NID, and IntensityHistogramCurveFit by validating the intermediated data such as GLCM matrix, NID matrix, and the fitted Gaussian curves. It is impossible to quantitatively validate them against *CGITA* because *CGITA* implementation is mainly for PET images and its ROI boundary handling is different from *IBEX*, *PINNACLE* TPS, and *ECLIPSE*.

At the time of this writing, *IBEX* has been used for two substantial projects<sup>11,15</sup> and is currently being used by around 35 researchers from different countries with CT (including contrast-enhanced CT, noncontrast-enhanced CT, cone beam CT, and 4D CT), PET, and MRI images. ROIs have been successfully imported from commercial and research software such as *ECLIPSE*, *PINNACLE*, *MIMvista*, *VELOCITY*, *SLICER*, and *MIRADA*. Researchers were able to use ROI editors to create new ROIs and modify the existing ROIs in *IBEX* ROI editor. Several researchers reported that *IBEX* is intuitive, powerful, and easy to use.

## 2.J. Distribution

Windows version 1.0 $\beta$  of *IBEX* is freely distributed. About 35 researchers around the world are using it and have contributed to the development of new preprocessing and feature extraction algorithms and review callback functions. The stand-alone version of *IBEX* without the requirement of a *MATLAB* license can be downloaded at [http://bit.ly/IBEX\\_MDAnderson](http://bit.ly/IBEX_MDAnderson). The source-code *IBEX* version requires installation of *MATLAB* and can be downloaded at <http://bit.ly/>

[IBEXSrc\\_MDAnderson](http://bit.ly/IBEX_MDAnderson) for free. Both versions of *IBEX* can be shipped via compact disc. *IBEX*-related documents can be found at [http://bit.ly/IBEX\\_Documentation](http://bit.ly/IBEX_Documentation).

An *IBEX* discussion group is available for users to post and answer any *IBEX*-related questions. Users can review the discussion threads at [https://groups.google.com/forum/#!forum/IBEX\\_users](https://groups.google.com/forum/#!forum/IBEX_users). Individuals can subscribe to the group by e-mailing [IBEX\\_users+subscribe@googlegroups.com](mailto:IBEX_users+subscribe@googlegroups.com) to obtain posting rights.

## 3. DISCUSSION

*IBEX* implemented the underlying modules and framework for radiomics and quantitative imaging analysis. *IBEX* serves as an open infrastructure software platform to accelerate collaborative work. Using *IBEX*, researchers can focus on their application and development of radiomics workflows without worrying about data consistency and review, algorithm reliability, and result reproducibility. The *IBEX* plug-in mechanism facilitates contribution of creative algorithms and implementation of customized requirements by users around the world.

Model development in quantitative imaging analysis is a major topic involving how to analyze and/or classify features. Many approaches to model development can be used, such as regression, principal component analysis, artificial neural networks, Bayesian networks, and support vector machines. Model development techniques can differ greatly depending on individual model applications. At this point, establishing a universal workflow for model development is difficult, so the current version of *IBEX* does not provide a tool for model development.

The *IBEX* developer studio is only available in the source-code version. This is because the stand-alone *MATLAB* program does not run unencrypted M-files. In other words, *MATLAB* does not allow mixing an encrypted M-file from the stand-alone version of *IBEX* with an unencrypted M-file from the *IBEX* developer studio. Developing the source code within the *MATLAB* environment and on the *IBEX* platform is always a good practice, as it enables advanced users to use the debugging and testing functionalities of both.

The *IBEX* database has a file-based structure and is organized in the same way in which *PINNACLE* native data storage is organized. Also, the *PINNACLE* native data format is used as the *IBEX* data format. As a result, *IBEX* data can be imported directly into the *PINNACLE* system. The *PINNACLE* data format basically has two parts: (1) the readable and modifiable ASCII header file describing the data and (2) the corresponding raw binary data. *PINNACLE* format data can be read quickly and efficiently, as a series of DICOM images is stored in one large portion of binary data. Users can use any text editor to open the ASCII *PINNACLE* header file to explore the data and modify the information as needed.

Although version 1.0 $\beta$  of *IBEX* has a radiomics infrastructure platform, we have been diligently working on the next version of *IBEX*, mainly focusing on improving the convenience and robustness of multi-institution, multidisciplinary collaborative research. Our near-term development goals include adding functions to do the following:

- Export intermediate data from IBEX for users to be able to check and use it for other research purposes.
- Archive completed projects including all necessary information so that project-related data can be restored or shared if reproduction or repeating of any analyses is needed.
- Add additional data importers (e.g., CERR format) as identified by the user network. Although use of DICOM is fairly standard for importing images and ROIs, many other formats can be used for images and/or delineated structures.
- Develop an extension for 3D SLICER to bridge 3D SLICER and IBEX.

#### 4. SUMMARY

We successfully implemented IBEX, an open infrastructure software platform that streamlines common radiomics workflow tasks. Its transparency, flexibility, and portability can greatly accelerate the pace of radiomics and its collaborative research and pave the way toward successful clinical translation. IBEX flexibly supports common radiomics workflow tasks such as multimodality imaging data import and review, development of feature extraction algorithms, model validation, and consistent data sharing among multiple institutions. On one hand, IBEX is self-contained and ready to use, with preimplemented typical data importers, image filters, and feature extraction algorithms. On the other hand, users can extend IBEX's functionality by plugging in new algorithms. IBEX also supports quality assurance for data and feature extraction algorithms. Image data, feature algorithms, and model formulas can be easily and consistently shared using IBEX for reproducibility purposes.

#### ACKNOWLEDGMENT

Conflicts of interest and sources of funding: Supported in part by a grant from the NCI (R03CA178495-01).

- <sup>a1</sup>Author to whom correspondence should be addressed. Electronic mail: LECourt@mdanderson.org; Telephone: 713-563-2546; Fax: 713-563-2479.
- <sup>1</sup>H. Y. Chen, S. L. Yu, C. H. Chen, G. C. Chang, C. Y. Chen, A. Yuan, C. L. Cheng, C. H. Wang, H. J. Terng, S. F. Kao, W. K. Chan, H. N. Li, C. C. Liu, S. Singh, W. J. Chen, J. J. Chen, and P. C. Yang, "A five-gene signature and clinical outcome in non-small-cell lung cancer," *N. Engl. J. Med.* **356**, 11–20 (2007).
- <sup>2</sup>E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," *Eur. J. Cancer* **45**, 228–247 (2009).
- <sup>3</sup>L. Fass, "Imaging and cancer: A review," *Mol. Oncol.* **2**, 115–152 (2008).
- <sup>4</sup>M. Machtay, F. Duan, B. A. Siegel, B. S. Snyder, J. J. Gorelick, J. S. Reddin, R. Munden, D. W. Johnson, L. H. Wilf, A. DeNittis, N. Sherwin, K. H. Cho, S. K. Kim, G. Videtic, D. R. Neumann, R. Komaki, H. Macapinlac, J. D. Bradley, and A. Alavi, "Prediction of survival by [18F]fluorodeoxyglucose positron emission tomography in patients with locally advanced non-small-cell lung cancer undergoing definitive chemoradiation therapy: Results of the ACRI 6668/RTOG 0235 trial," *J. Clin. Oncol.* **31**, 3823–3830 (2013).
- <sup>5</sup>D. J. Raz, M. R. Ray, J. Y. Kim, B. He, M. Taron, M. Skrzypski, M. Segal, D. R. Gandara, R. Rosell, and D. M. Jablons, "A multigene assay is prognostic

- of survival in patients with early-stage lung adenocarcinoma," *Clin. Cancer Res.* **14**, 5565–5570 (2008).
- <sup>6</sup>O. S. Al-Kadi and D. Watson, "Texture analysis of aggressive and nonaggressive lung tumor CE CT images," *IEEE Trans. Biomed. Eng.* **55**, 1822–1830 (2008).
- <sup>7</sup>S. Basu, "Developing predictive models for lung tumor analysis," M.S. thesis, University of South Florida, 2012.
- <sup>8</sup>A. R. Cunliffe, H. A. Al-Hallaq, Z. E. Labby, C. A. Pelizzari, C. Straus, W. F. Sensakovic, M. Ludwig, and S. G. Armato, "Lung texture in serial thoracic CT scans: Assessment of change introduced by image registration," *Med. Phys.* **39**, 4679–4690 (2012).
- <sup>9</sup>A. R. Cunliffe, S. G. Armato III, X. M. Fei, R. E. Tuohy, and H. A. Al-Hallaq, "Lung texture in serial thoracic CT scans: Registration-based methods to compare anatomically matched regions," *Med. Phys.* **40**, 061906 (9pp.) (2013).
- <sup>10</sup>A. R. Cunliffe, S. G. Armato, C. Straus, R. Malik, and H. A. Al-Hallaq, "Lung texture in serial thoracic CT scans: Correlation with radiologist-defined severity of acute changes following radiation therapy," *Phys. Med. Biol.* **59**, 5387–5398 (2014).
- <sup>11</sup>B. Ganeshan, S. Abaleke, R. C. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: Initial evidence for a relationship with tumour glucose metabolism and stage," *Cancer Imaging* **10**, 137–143 (2010).
- <sup>12</sup>B. Ganeshan, V. Goh, H. C. Mandeville, Q. S. Ng, P. J. Hoskin, and K. A. Miles, "Non-small cell lung cancer: Histopathologic correlates for texture parameters at CT," *Radiology* **266**, 326–336 (2013).
- <sup>13</sup>B. Ganeshan, E. Panayiotou, K. Burnand, S. Dizdarevic, and K. Miles, "Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: A potential marker of survival," *Eur. Radiol.* **22**, 796–802 (2012).
- <sup>14</sup>L. A. Hunter, S. Krafft, F. Stingo, H. Choi, M. K. Martel, S. F. Kry, and L. E. Court, "High quality machine-robust image features: Identification in nonsmall cell lung cancer computed tomography images," *Med. Phys.* **40**, 121916 (12pp.) (2013).
- <sup>15</sup>M. Ravanelli, D. Farina, M. Morassi, E. Roca, G. Cavalleri, G. Tassi, and R. Maroldi, "Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: Prediction of the response to the first-line chemotherapy," *Eur. Radiol.* **23**, 3450–3455 (2013).
- <sup>16</sup>E. Segal, C. B. Sirlin, C. Ooi, A. S. Adler, J. Gollub, X. Chen, B. K. Chan, G. R. Matcuk, C. T. Barry, H. Y. Chang, and M. D. Kuo, "Decoding global gene expression programs in liver cancer by noninvasive imaging," *Nat. Biotechnol.* **25**, 675–680 (2007).
- <sup>17</sup>H. Tan, T. Liu, Y. Wu, J. Thacker, R. Shenkar, A. G. Mikati, C. Shi, C. Dykstra, Y. Wang, P. V. Prasad, R. R. Edelman, and I. A. Awad, "Evaluation of iron content in human cerebral cavernous malformation using quantitative susceptibility mapping," *Invest. Radiol.* **49**, 498–504 (2014).
- <sup>18</sup>T. Win, K. A. Miles, S. M. Janes, B. Ganeshan, M. Shastry, R. Endozo, M. Meagher, R. I. Shortman, S. Wan, I. Kayani, P. J. Ell, and A. M. Groves, "Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer," *Clin. Cancer Res.* **19**, 3591–3599 (2013).
- <sup>19</sup>D. V. Fried, S. L. Tucker, S. Zhou, Z. Liao, O. Mawlawi, G. Ibbott, and L. E. Court, "Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer," *Int. J. Radiat. Oncol., Biol., Phys.* **90**, 834–842 (2014).
- <sup>20</sup>P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer* **48**, 441–446 (2012).
- <sup>21</sup>V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, and R. J. Gillies, "Radiomics: The process and the challenges," *Magn. Reson. Imaging* **30**, 1234–1248 (2012).
- <sup>22</sup>A. Jackson, J. P. O'Connor, G. J. Parker, and G. C. Jayson, "Imaging tumor vascular heterogeneity and angiogenesis using dynamic contrast-enhanced magnetic resonance imaging," *Clin. Cancer Res.* **13**, 3449–3459 (2007).
- <sup>23</sup>C. J. Rose, S. Mills, J. P. O'Connor, G. A. Buonaccorsi, C. Roberts, Y. Watson, B. Whitcher, G. Jayson, A. Jackson, and G. J. Parker, "Quantifying heterogeneity in dynamic contrast-enhanced MRI parameter maps," *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Springer, Verlag/Berlin/Heidelberg (NY), 2007), Vol. 10, pp. 376–384.

- <sup>24</sup>P. Gibbs and L. W. Turnbull, "Textural analysis of contrast-enhanced MR images of the breast," *Magn. Reson. Med.* **50**, 92–98 (2003).
- <sup>25</sup>H. C. Canuto, C. McLachlan, M. I. Kettunen, M. Velic, A. S. Krishnan, A. A. Neves, M. de Backer, D. E. Hu, M. P. Hobson, and K. M. Brindle, "Characterization of image heterogeneity using 2D Minkowski functionals increases the sensitivity of detection of a targeted MRI contrast agent," *Magn. Reson. Med.* **61**, 1218–1224 (2009).
- <sup>26</sup>J. O. Deasy, A. I. Blanco, and V. H. Clark, "CERR: A computational environment for radiotherapy research," *Med. Phys.* **30**, 979–985 (2003).
- <sup>27</sup>Y. H. Fang, C. Y. Lin, M. J. Shih, H. M. Wang, T. Y. Ho, C. T. Liao, and T. C. Yen, "Development and evaluation of an open-source software package 'CGITA' for quantifying tumor heterogeneity with molecular images," *BioMed Res. Int.* **2014**, 248505 (2014).
- <sup>28</sup>P. M. Szczypinski, M. Strzelecki, A. Materka, and A. Klepaczko, "MaZda—A software package for image texture analysis," *Comput. Methods Programs Biomed.* **94**, 66–76 (2009).
- <sup>29</sup>G. E. Krasner and S. T. Pope, "A cookbook for using the model-view controller user interface paradigm in Smalltalk-80," *J. Object Oriented Program.* **1**, 26–49 (1988).
- <sup>30</sup>T. Xie, K. Taneja, S. Kale, and D. Marinov, "Towards a framework for differential unit testing of object-oriented programs," in *Proceedings of the Second International Workshop on Automation of Software Test* (IEEE Computer Society, Washington, DC, 2007), p. 5.
- <sup>31</sup>P. Mildemberger, M. Eichelberg, and E. Martin, "Introduction to the DICOM standard," *Eur. Radiol.* **12**, 920–927 (2002).
- <sup>32</sup>R. M. Haralick, K. Shanmuga, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.* **3**, 610–621 (1973).
- <sup>33</sup>H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Cavalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebbers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.* **5**, 4006 (2014).
- <sup>34</sup>M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Trans. Syst., Man, Cybern.* **19**, 1264–1274 (1989).
- <sup>35</sup>X. O. Tang, "Texture information in run-length matrices," *IEEE Trans. Image Process.* **7**, 1602–1609 (1998).
- <sup>36</sup>M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graphics Image Process.* **4**, 172–179 (1975).