

COMPLEMENTARY FEATURES FOR RADIOMIC ANALYSIS OF MALIGNANT AND BENIGN MEDIASTINAL LYMPH NODES

Tuan D. Pham

Department of Biomedical Engineering
Linköping University
Linköping 58183, Sweden

ABSTRACT

The importance of developing effective strategies for investigating mediastinal lymph-node metastases in non-small cell lung cancers is increasingly emphasized. It is because the precise detection of this metastatic disease is critical for optimal surgical intervention and treatment for patients with lung cancer. Existing medical image analysis is of limited power for mediastinal lymph-node staging on computed tomography (CT). Motivated by the radiomics hypothesis, this paper explored deep-learning, texture features and their combinations to ascertain subtle difference between malignant and benign mediastinal lymph nodes on CT. The radiomics-based results are found to be promising for differentiating malignant from benign mediastinal lymph nodes of patients with lung cancer.

Index Terms— Lung metastasis, benign, computed tomography, radiomics, deep learning, texture.

I. INTRODUCTION

Advanced image analysis of CT and PET/CT for improving the assessment of lymph-node metastasis from patients with lung cancer is still in early development. CT-based texture analysis is currently the most popular approach for differentiating between benign and malignant mediastinal lymph nodes [1], [2], [3], where texture analysis results are found promising for improving sensitivity and specificity of the lymph-node metastasis, and other medical imaging problems [4], [5]. After the emergence of the hypothesis of radiomics [6] in oncology, which suggests the existence of statistical but hidden information in radiological images, scientific explorations of radiomic features have been done to utilize this source of information that is still to be much further investigated for medical prognosis and prediction of clinical outcomes for cancer patients [7]-[9]. Radiomics refers to the high-throughput extraction and analysis of large amounts of quantitative features from medical images obtained with CT, PET or magnetic resonance imaging (MRI). The mining of radiomic data is potential for building descriptive and predictive models that relate non-invasive features to phenotype signatures of cancer. Indeed, the association between radiographic features and mediastinal lymph-node metastasis has been confirmed [10]. The radiomics hypothesis therefore motivates a need for creating automated

and reproducible analysis methodologies to extract more information from image-based features that are expected to facilitate better clinical decision making, particularly in the care of patients with cancer at low cost [7], [8], [9].

One of the most attractive methods for mining hidden image features of heterogeneous objects is the deep machine learning or deep-learning approach that attempts to discover hierarchical patterns in data structure by using cascaded neural networks with multiple processing units [11], [12]. Because texture has been recognized as a popular approach for medical image analysis, this paper further explores the complimentary combination of deep-learning and texture features in order to improve the discrimination of the benign and malignant lymph-node regions. Taking both the radiomics hypothesis and deep learning into account, this paper reports a study in the extraction of hidden features in CT images of mediastinal lymph nodes to discriminate benign and malignant samples by utilizing these two emerging concepts. The rest of this paper is organized as follows. Section II includes methods that describe clinical data of lung-cancer patients, design of deep learning in neural networks, and the semivariogram in geostatistics. Section III presents experimental results and discussion, including comparisons of the results with those obtained from similar studies. Finally, Section IV is the conclusion of the research finding.

II. METHODS

II-A. CT Data

The image data, which was described earlier [13], were obtained with inclusion criteria comprised biopsy-proven primary lung malignancy with pathological mediastinal nodal staging and unenhanced CT of the thorax. A total of 148 consecutive patients were included (93 men, 55 women, 36-84 years of age). There were 105 adenocarcinomas, 28 squamous cell carcinomas, 6 adenosquamous cell carcinomas, 5 large cell carcinomas, 1 small cell lung cancer, 3 pleomorphic carcinomas. Histological analysis were obtained in surgical resection (148 patients). Most biopsied lymph nodes were at stations 7 and 4R (no. 4 right lymph nodes). Figure 1 shows some regions of benign and malignant mediastinal lymph nodes.

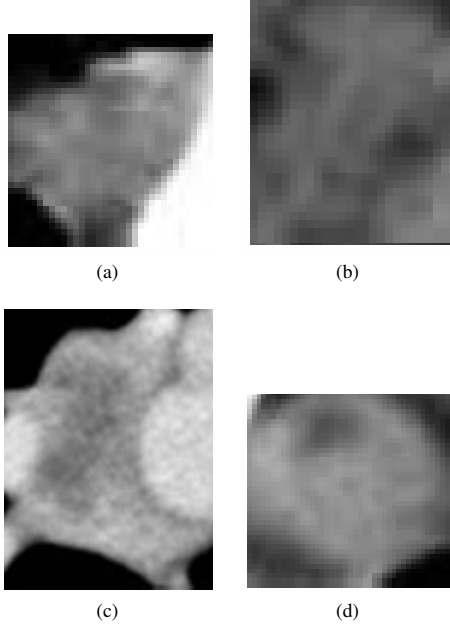


Fig. 1: CT regions of lung lymph nodes: (a)-(b): benign, and (c)-(d): malignant.

II-B. Deep learning for feature mining

The discovery of hidden and effective features of an object can be obtained by deep learning with the implementation of an autoencoder that is an unsupervised neural network. An autoencoder consists of two components: an encoder and a decoder. An encoder in the hidden layer k maps the input $\mathbf{x} \in \mathcal{R}^{D_{\mathbf{x}}}$ to another representation $\mathbf{a}^k \in \mathcal{R}^{D^k}$ by means of a transfer function f , and can be expressed as follows

$$\mathbf{a}^k = f(\mathbf{x}) = f(\mathbf{W}^k \mathbf{x} + \mathbf{b}^k), \quad (1)$$

where $\mathbf{W}^k \in \mathcal{R}^{D^k \times D_{\mathbf{x}}}$ is a weight matrix, and $\mathbf{b}^k \in \mathcal{R}^{D^k}$ is a bias vector.

Using a transfer function g , a decoder learns to reconstruct the original input \mathbf{x} , denoted as \mathbf{y} , and is defined as

$$\mathbf{y} = g[\mathbf{a}^{(k+1)}] = g[\mathbf{W}^{(k+1)} \mathbf{x} + \mathbf{b}^{(k+1)}]. \quad (2)$$

Figure 2 shows the structure of the autoencoder implemented in this study.

The optimal learning process of the autoencoder is performed by minimizing the following loss function \mathcal{L} [11]:

$$\mathcal{L} = L(\mathbf{x}, \mathbf{y}) + \lambda_s \Omega(\mathbf{a}) + \lambda_r \Omega(\mathbf{w}), \quad (3)$$

where $L(\mathbf{x}, \mathbf{y})$ is a loss function such as the mean squared error that imposes a cost for the difference between \mathbf{x} and its reconstructed signal \mathbf{y} , $\Omega(\mathbf{a})$ is a sparsity penalty, λ_s is the coefficient for the sparse penalty, $\Omega(\mathbf{w})$ is the L_2 regularizer and λ_r its coefficient. These terms are mathematically defined as follows.

$$\Omega(\mathbf{a}) = \sum_{i=1}^n KLD(p || \tilde{p}_i), \quad (4)$$

where $KLD(p || \tilde{p}_i)$ is the Kullback-Leibler divergence, which is used as a measure of the difference between a sparsity parameter p and average activation value \tilde{p}_i , n is the number of neurons in the hidden layer. The value for p is close to zero, typically taken as 0.05 [11], and \tilde{p}_i is defined with rewriting a_i^k as $a_i^k(x_j)$ to explicitly express the activation of a neuron i in hidden unit k when the network is given a specific input x_j as follows

$$\tilde{p}_i = \frac{1}{N} \sum_{j=1}^N a_i^k(x_j), \quad (5)$$

where N is the total number of training samples.

The L_2 regularizer attempts to add a regularization term on the weights to the loss function:

$$\Omega(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^L \sum_{j=1}^N \sum_{i=1}^V (w_{ji}^k)^2 \quad (6)$$

where L , N , and V are the numbers of hidden layers, training samples, and training-data variables, respectively.

II-C. Semivariogram for texture extraction

Let $f(x_i)$, $i = 1, 2, \dots, S$, be image intensity values of an image of size S pixels as the product of its numbers of rows and columns (sampling space), the semivariogram of an image can be expressed as follows [14]

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [f(x_i) - f(x_i + h)]^2, \quad (7)$$

where $N(h)$ is the number of pairs of variables separated by distance or lag h , which is taken as the number of pixels.

The above expression is based on the assumption that the spatial autocorrelation structure is isotropic, which implies the semi-variogram depends only on the magnitude of the lag (h). When the spatial autocorrelation pattern changes in different directions in the sampling space, an anisotropic semi-variogram can be of more appropriate use. Geometric and zonal properties of anisotropy have been introduced to model the semivariogram [15]. The geometric anisotropy is imposed when the range of the semivariogram varies in different directions. The zonal anisotropy is a phenomenon when the range and sill of the semivariogram are not constant.

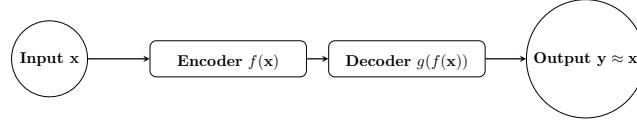


Fig. 2: Structure of autoencoder that maps input x to output y through encoder f and decoder g , where feature set is generated by encoder. Second set of features can be generated by passing first feature set through encoder of second autoencoder.

III. RESULTS AND DISCUSSION

To extract deep-learning features of the metastatic and benign lymph nodes from the patients with lung cancer, the images were scaled to the same size in order to enable the unsupervised training of the sparse autoencoders and autoencoders at different network layers. Based on recent findings [2], [13] reporting that benign and malignant lymph nodes consist of different characteristics of ground tissues (parenchyma) around them, instead of carrying the image segmentation of the lymph nodes, regions of lymph nodes including the ground tissues, were manually selected in this study. Because the image sizes of the regions of interest containing malignant lymph nodes vary between 16×24 and 84×85 pixels, and benign lymph nodes between 26×18 and 122×112 pixels; the image size of 50×50 pixels, which is an approximately average size, was selected for the rescaling of the images.

The deep-learning features extracted from the images were obtained by training two autoencoders, where the first feature set obtained from the first autoencoder was used as the input to the second autoencoder that produced the second or final set of features. The adopted transfer function for both the encoder and decoder is the logistic sigmoid function. For deep-learning training, given the image size of $50 \times 50 = 2500$ pixels, the size of the units of the hidden layer for the first autoencoder was chosen to be 200 (much smaller than 2500), maximum number of epochs = 400, $\lambda_r = 0.004$ (coefficient for the L_2 regularization term), $\lambda_s = 4$ (coefficient for the sparsity regularization term), and sparsity proportion parameter = 0.15 (low sparsity proportion indicates higher degree of encoder sparsity). The number of units of the hidden layer for the second autoencoder was decreased to 100 (half of the size of the first autoencoder), $\lambda_r = 0.002$, $\lambda_s = 4$, and sparsity proportion parameter = 0.1. These parameter specifications mean that the original vectors of 2500 dimensions in the first encoder were reduced to 200 dimensions; and after passing through the second encoder, the vectors of 200 dimensions were further reduced to 100-dimensional vectors that were finally used as deep-learning features for differentiating between the benign and malignant lymph nodes.

The deep-learning (DL), gray-level co-occurrence matrix (GLCM) [16], local binary patterns (LBP) [17], semi-variogram (SV) features of the benign and malignant lymph nodes were fitted to a logistic regression model using the binomial distribution to compute the receiver operating char-

acteristic (ROC) curve, which is a plot of the true positive rate also known as sensitivity (SEN) or recall, versus the false positive rate calculated as 1-specificity, where specificity (SPE) is the true negative rate, to show the performance of a binary classification with varying thresholds. For the GLCM, 20 GLCM-based features were used [16], [18], [19]: entropy, energy, correlation, contrast, sum of squares (variance), sum average, sum variance, sum entropy, difference variance, difference entropy, information measures of correlation, autocorrelation, dissimilarity, homogeneity, cluster prominence, cluster shade, maximum probability, inverse difference, inverse difference normalized, and inverse difference moment normalized. The LBP feature vectors were extracted using 8 neighbors to compute the local binary pattern for each pixel in an image, the set of neighbors was selected from a circularly symmetric pattern around each pixel with radius = 1, and the L_2 normalization was applied to the LBP histograms. The SV vector features were extracted with 30 lags in horizontal and vertical directions of the images.

Table I: ROC and other measures obtained from logistic regression analysis of texture and deep learning features.

Feature type	AUC	SEN (%)	SPE (%)	PPV	NPV	LR+
GLCM	0.7165	75.19	60.14	65	72	2
LBP	0.8199	80.45	74.64	75	80	3
SV	0.6207	59.40	64.49	62	62	2
DL	0.9226	89.47	81.16	82	89	5
DL+LBP	0.9412	86.47	88.41	88	87	7
DL+GLCM	0.8125	84.96	77.54	78	84	4
DL+SV	0.9464	83.46	93.48	93	85	13
DL+SV+LBP	0.9402	84.96	88.41	88	86	7

Table II: Comparisons of ROC characteristics.

Features	# Patients	AUC	SEN (%)	SPE (%)
SVV_{max} [21]	219	0.79	74	73
GLCM+GLNU+RLNU+shape [1]	43	0.87	81	80
IFH [3]	29	0.83	53	97
DL	148	0.92	89	81
DL+SV	148	0.95	83	94

In addition to sensitivity and specificity, positive predictive value (PPV), negative predictive value (NPV), and positive likelihood ratio (LR+) are also popular quantitative measures for assessing a clinical diagnostic test [20]. The PPV aims to provide a degree of confidence to clinicians about the patient has the disease given that the test is positive, whereas the NPV answers how likely the patient does not have the disease given that the test is negative. The

Table III: Ten-fold cross-validation results.

Feature type	AUC	SEN (%)	SPE (%)
GLCM	0.6475	61.79	61.51
LBP	0.6862	62.50	73.48
SV	0.5804	58.46	64.18
DL	0.7260	71.29	68.04
DL+LBP	0.7537	63.65	84.71
DL+GLCM	0.6576	60.49	71.34
DL+SV	0.7595	68.08	75.27
DL+SV+LBP	0.7512	68.01	73.35

LR+ tells how much to increase the probability of disease of a patient whose test is positive with respect to a patient whose test is negative (LR+ = 1 indicates that there is no change in the probability of disease with the test result).

Table I shows the AUC, SEN, SPE, PPV, NPV, and LR+ values obtained from the logistic regression analysis of the gray-level co-occurrence matrix (GLCM), local binary patterns (LBP), semivariogram (SV), and deep learning (DL) features. For individual features, the DL features provide the best results with AUC = 0.9226, SEN = 89.47%, and SPE = 81.16%. The AUC obtained from the LBP is higher than those obtained from the SV and GLCM. The AUC obtained from the SV is the smallest among the other three individual features. Regarding the feature combination, it can be clearly seen that the combined features of DL and SV are most complementary, although the AUC of the latter is smallest, resulting in the largest AUC of 0.95, with SEN = 83.45%, and SPE = 93.48%. The AUC (0.9402) obtained from the combination of the DL and LBP is slightly lower than the AUC obtain from the combination of the DL and SV. The combined DL and GLCM improves the AUC (0.8125) obtained from the individual GLCM (0.7165), but it is smaller than the AUC (0.9226) obtained from the individual DL.

Regarding the PPV, NPV, and LR+ obtained from the four feature types, the DL features provide measures that give strongest evidence to confirm or refute the diagnosis of lung metastasis, with PPV = 82% and NPV = 89%. The LBP gives second strongest results with PPV = 75% and NPV = 80%. The GLCM gives stronger evidence for NPV (72%) than the NPV obtained from the SV (65%), while the PPV obtained from the SV (73%) is larger than that obtained from the GLCM (65%). An LR+ value in the range from 5 to 10 indicates a moderate conclusive increase in the likelihood of disease, whereas an LR+ between 2 and 4 implies a small increase in the likelihood of disease. Here, the DL feature provides an LR+ of 5, while the other three individual features are among 2 and 3. However, the combination of the DL and SV resulted in a strikingly larger value of the LR+, which is 13, where LR+ > 10 is known to be an excellent measure for clinical tests.

Table II shows the ROC characteristics obtained from the logistic regression analysis of benign and malignant mediastinal lymph nodes on CT, where PET/CT data and maximum standard uptake value SUV_{max} were used in [21], GLCM, gray-level nonuniformity (GLNU) and run-length nonuniformity (RLNU) texture, and shape features used in

[1], and image filtration-histogram (IFH) texture used in [3]. The results obtained from the extraction of DL features and their combination with the SV features proposed in this study are superior to all those obtained from the other three studies recently reported in literature in terms of the area under curve, sensitivity, specificity, positive predictive value, and negative predictive value. It should be pointed out that although the specificity provided in [3], which is 97%, is higher than the specificity provided by the individual DL (81%) and combined DL and SV features (94%); the DL-based sensitivity (89%) and combined DL-SV sensitivity (83%) are much higher than the other sensitivity (53%). The combination of DL and SV features provides the largest AUC, which is 0.95, among all other results.

The performance of the logistic regression analysis of the benign and malignant mediastinal lymph nodes on CT was validated using the 10-fold cross-validation, where the validated results of the area under curve (AUC), sensitivity (SEN), and specificity (SPE) obtained from the individual and combined features are presented in Table III. Both sets of results suggest the high performance of the DL (largest AUC) among GLCM, LBP, and SV features, and its complementary combinations with the three textures.

Here, each image size is of $50 \times 50 = 2500$ pixels, by setting the number of hidden layers in the first autoencoder to 100, which was much less than the image size, the neural network was forced to learn a sparse representation of the input images to become its compressed version. The number of units of the hidden layer for the second autoencoder was then reduced to 50, once again the network was to learn a more compressed representation of the input images in order to extract more informative features by discarding redundant information of the raw input. If the input has a random pattern, then it would be a challenge for the sparse coding task of the autoencoders. However, according to the hypothesis of radiomics, relationships between imaging phenotypes of disease and control exist, this sparse representation of learning a lower dimension enables the discovery of such relationships to some certain extent. It should be pointed out that, in this study, some images were downsapmled and some upsampled to fit the size of 50×50 pixels for the uniformed training of the autoencoders. When the number of hidden units were larger than the number of pixels, the discovery of useful features was obtained by setting the sparsity penalty on the hidden units of the autoencoders. Such cases of feature discovery using deep learning have also been reported in literature [22].

IV. CONCLUSION

The experimental results suggest the complementary combination of CT hierarchical and textural information in thoracic lymph nodes of patients with lung cancer, which can achieve higher classification accuracy. Research into mathematical methods for fusion of image features as well as applications of other deep-learning models and advanced machine-learning methods for pattern classification are worthy of further investigation to advance the state-of-the art in image analysis for lung-cancer diagnosis and prognosis [13].

V. REFERENCES

- [1] H. Bayanati, R.E. Thornhill, C.A. Souza, V. Sethi-Virmani, A. Gupta, D. Maziak, K. Amjadi, C. Dennie, "Quantitative CT texture and shape analysis: Can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer?," *Eur Radiol.*, vol. 25, pp. 480-487, 2015.
- [2] S.K.N. Dilger, J. Uthoff, A. Judisch, E. Hammond, S.L. Mott, B.J. Smith, J.D. Newell, E.A. Hoffman, J.C. Sieren, "Improved pulmonary nodule classification utilizing quantitative lung parenchyma features," *J. Medical Imaging*, vol. 2, 041004, 2015.
- [3] M.B. Andersen, S.W. Harders, B. Ganeshan, J. Thygesen, H.H.T. Madsen, F. Rasmussen, "CT texture analysis can help differentiate between malignant and benign lymph nodes in the mediastinum in patients suspected for lung cancer," *Acta Radiologica*, vol. 57, pp. 669-676, 2016.
- [4] T.D. Pham, D.T.P. Le, J. Xu, D.T. Nguyen, R.G. Martindale, C.W. Deveney, "Personalized identification of abdominal wall hernia meshes on computed tomography," *Computer Methods and Programs in Biomedicine*, vol. 113, pp. 153-161, 2014.
- [5] T. Tsunoyama, T.D. Pham, T. Fujita, T. Sakamoto, "Identification of intestinal wall abnormalities and ischemia by modeling spatial uncertainty in computed tomography imaging findings," *Computer Methods and Programs in Biomedicine*, vol. 117, pp. 30-39, 2014.
- [6] 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, R.J. Gillies, "Radiomics: the process and the challenges," *Magnetic Resonance Imaging*, vol. 30, pp. 1234-1248, 2012.
- [7] R.J. Gillies, P.E. Kinahan, H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, pp. 563-577, 2016.
- [8] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G.P.M. van Stiphout, P. Granton, C.M.L. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J.W.L. Aerts, "Radiomics: extracting more information from medical images using advanced feature analysis," *Eur J Cancer*, vol. 48, pp. 441-446, 2012.
- [9] H.J.W.L. Aerts, Emmanuel Rios Velazquez, R.T.H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haihe-Kains, D. Rietveld, F. Hoebbers, M.M. Rietbergen, C.R. Leemans, A. Dekker, J. Quackenbush, R.J. Gillies, P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, 2014. DOI: 10.1038/ncomms5006.
- [10] Y. Xia, B. Zhang, H. Zhang, W. Li, K.P. Wang, H. Shen, "Evaluation of lymph node metastasis in lung cancer: who is the chief justice?," *J Thorac Dis*, vol. 7, pp. S231-S237, 2015.
- [11] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*. MIT Press (in preparation), 2016, <http://www.deeplearningbook.org> (last accessed 01 August 2016).
- [12] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT Press, 2012.
- [13] T.D. Pham, Y. Watanabe, M. Higuchi, H. Suzuki, "Texture analysis and synthesis of malignant and benign mediastinal lymph nodes in patients with lung cancer on computed tomography," *Scientific Reports*, vol. 7, 43209, 2017. DOI: 10.1038/srep43209.
- [14] T.D. Pham, "The semi-variogram and spectral distortion measures for image texture retrieval," *IEEE Trans Image Processing*, vol. 25, pp. 1556-1565, 2016.
- [15] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications*. Berlin: Springer-Verlag, 2003.
- [16] R.M. Haralick, K. Shanmugam, I. Dinstein, "Textural features of image classification," *IEEE Trans Systems, Man and Cybernetics*, vol. 3, pp. 610-621, 1973.
- [17] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray scale and rotation invariant texture classification with local binary patterns," *IEEE Trans Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971-987, 2002.
- [18] L. Soh, C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans Geoscience and Remote Sensing*, vol. 37, pp. 780-795, 1999.
- [19] D.A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Can. J. Remote Sensing*, vol. 28, pp. 45-62, 2002.
- [20] A.G. Lalkhen, A. McCluskey, "Clinical tests: sensitivity and specificity," *Contin Educ Anaesth Crit Care Pain*, vol. 8, pp. 221-223, 2008.
- [21] S. Li, Q. Zheng, Y. Ma, Y. Wang, Y. Feng, B. Zhao, Y. Yang, "Implications of false negative and false positive diagnosis in lymph node staging of NSCLC by means of 18F-FDG PET/CT," *PLoS ONE*, vol. 8, e78552, 2013.
- [22] UFLDL (Unsupervised Feature Learning and Deep Learning) Tutorial, *Autoencoders*, <http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders> (last accessed 01 August 2016).