Dimensionality Reduction of Mass Spectrometry Imaging Data using Autoencoders

Spencer A. Thomas *, Alan M. Race*,
Rory T. Steven*, Ian S. Gilmore*, Josephine Bunch*†
*National Centre of Excellence in Mass Spectrometry Imaging (NiCE-MSI)
National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK
†School of Pharmacy, University of Nottingham, University Park, Nottingham, UK
Email: {spencer.thomas; alan.race; rory.steven; ian.gilmore; josephine.bunch}@npl.co.uk

***Abstract*—The use of mass spectrometry imaging (MSI) techniques has become a powerful tool in the fields of biology, pharmacology and healthcare. Next generation experimental techniques are able to generate 100s of gigabytes of data from a single image acquisition and thus require advanced algorithms in order to analyse these data. At present, analytical work-flows begin with pre-processing of the data to reduce its size. However, the pre-processed data is also high in dimensionality and requires reduction techniques in order to analyse the data. At present, mostly linear dimensionality reduction techniques are used for hyper-spectral data. Here we successfully apply an autoencoder to MSI data with over 165,000 pixels and more than 7,000 spectral channels reducing it into a few core features. Our unsupervised method provides the MSI community with an effective non-linear dimensionality reduction technique which includes the mapping to and from the reduced dimensional space. This method has added benefits over methods such as PCA by removing the need to select meaningful features from the entire list of components, reducing subjectivity and significant human interaction from the analysis.**

I. Introduction

Determining the structure and distribution of unlabelled molecules within a sample is an area of interest in material, pharmaceutical, biological, healthcare, and forensic sciences. Mass spectrometry (MS) techniques are able to determine the relative abundance of these molecules and thus the composition of a material. MS now includes a wide range of methods, e.g. matrix-assisted laser desorption/ionisation (MALDI) [1], desorption electrospray ionization (DESI) [2] and secondary ionisation mass spectrometry (SIMS) [3]. In all cases molecules are desorbed and ionised from a surface and the resulting ions are separated according to their mass-to-charge ratio (*m/z*) in a mass analyser. The resulting intensity spectra, referred to as mass spectra, contain peaks that characterise particular compounds, or fragments of them, in the material.

Mass spectrometry imaging (MSI) techniques provides a powerful means of imaging the chemical content of a sample. That is, each pixel in a MS image corresponds to a mass spectrum; a 3D object where the x and y dimensions relate to spatial location, and the z dimension corresponds to the spectrum of *m/z* values, an illustration of this is given in Fig. 1. Using these data it is possible to build false colour images of a sample based on the intensity at each pixel for a particular *m/z*, i.e *x, y* slices along the z dimension in Fig. 1. These hyper-spectral data enable the determination of both the location and distribution of a particular compound in a sample. MSI can therefore provide both qualitative and sometimes quantitative differences in the chemical composition of a sample under different conditions. The ability to capture material composition and distribution has made MSI techniques extremely useful in analysing biological samples as they can identify drugs [4], [5] lipids [6], [7] peptides [8] proteins [9] and metabolites [4], [5] from many different tissues.

One significant, and growing, issue in the area of MSI is the size of the data. Depending on spatial and mass resolution, which can vary with not only techniques but also instrumentation, MSI data can range from gigabytes to a few terabytes for large image areas [10]. Analysis of such large data becomes problematic and rapidly prohibitive for computational techniques that require data to be stored in memory. Owing to the ever increasing data size in MSI, there is a growing need for collaborations with mathematicians and computer scientists. Recently the community has made important strives forward in the analytics of large datasets. For example, open source data repositories [11], studies of replicate datasets [12], the development of visualisation tools [13], [14] and concepts [15], and new scalable algorithms [16], [17]. Despite this progress, larger efforts are required to tackle the increasing data issues in MSI. These problems require multi-disciplinary solutions between physical and life sciences, with a focus on large-scale memory-efficient algorithms for data analysis.

This paper investigates the use of autoencoders for unsupervised non-linear dimensionality reduction of MSI data. We demonstrate how the autoencoder can extract the core features of a MALDI dataset in to a single hidden node that compares well to annotated reference material. The rest of the paper is organised as follows: Section II outlines in brief the MSI technique used to acquire the data in this work, and an overview of the analytical tools currently used by the MSI community. In Section III we describe how we apply the autoencoder to MSI data. Section IV introduces the dataset used in this work, with the results presented in Section V. Finally we conclude our investigation in Section VI.

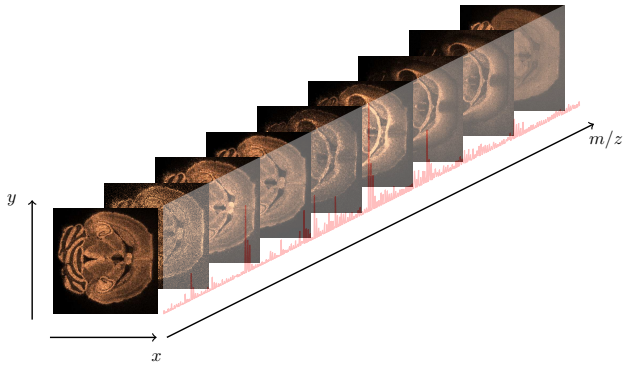


Fig. 1. Hyper-spectral mass spectrometry image data. So-called ion images are created by taking pixel intensity values at specific *m/z* values providing localisation and distribution of chemical species.

II. Matrix assisted laser desorption/ionisation (MALDI) mass spectrometry imaging

MALDI is a semi-quantitative MS technique can be used to generate ions of small and large molecules directly from complex samples. An incoming laser causes ablation and desorption of material from the surface of a sample. The desorbed material then forms positive and negative ions in the gas of ablated material, which are passed to a mass analyser to generate a mass spectrum. As both positive and negative ions are produced during the ionisation process, MALDI experiments can generate a mass spectrum based on the negative or the positive ions.

Typical samples are prepared into sections a few cm^2 in area and are 5-20 μm thick. The sample is then covered by a thin layer compound, typically a low molecular weight organic acid known as a 'matrix'. This compound is designed to increase absorption of the incoming laser energy, aid desorption / ionisation of the sample, and ultimately, increase the ion yield detected by the mass analyser.

When using MALDI to acquire an image of a sample, as with other MSI techniques, a mass spectrum is generated from each pixel location. This enables the spatially resolved detection of large in-tact molecules within the sample. Typical data acquisition is a few hours for a MALDI image, though will vary with image size and instrumentation.

A. Current Mass Spectrometry Imaging Analysis

Mass spectra are typically recorded over a large range of mass to charge ratios relative to the spectral resolving power of the instrument. As a result the raw data are typically sparse with narrow peaks containing both Poison and Gaussian noise components, as illustrated by the *m/z* dimension in Fig. 1. Preprocessing techniques exploit the sparsity of the data and reduce it to typically a few thousand *m/z* channels. This method is referred to as 'peak picking' where, due to the complexity of the data and the large number of spectra, several methods are used. Common peak picking work-flows are based on the mean spectrum for all pixels [18], removing peaks that appear in less than 1% of pixels [19], or based on a region of interest [20].

For MSI, even the *peak picked* data are complex and high in dimensionality. For example, a MS image with 100x100 pixels can contain 5,000-10,000 *m/z* values (and corresponding intensities) after peak picking for each pixel. This results in 100,000,000 values and file sizes of 10s of gigabytes. Therefore, the first process in most MSI work flows is an unsupervised dimensionality reduction technique. The most popular method is through principal component analysis (PCA) as it is able to separate features in the data based on decreasing variance. The largest source of variance, captured in the first few components, typically separates the data in to sample and substrate.

However, there are several issues with PCA. Firstly, it is a linear method. This is problematic as MSI data are complex due to the presence of non-linearities in the ionisation process and also the phenomenon of enhancement and suppression of detected ion intensity caused by the surrounding material [21]. Therefore, analytical methods that can account for non-linear behaviour are essential to better understand MSI data. Secondly, PCA separates the data in to orthogonal components of decreasing variance, and there is no way of knowing *a priori* how many components are required to separate the noise and the sample. This issue is compounded as the number of components that separate the sample and substrate varies between experiments. Moreover, as the number of components is related to the dimensionality of the dataset, many of the principal components are trivial, and there is no automated method of determining which ones provide insight. It is also unclear how many components are needed to reconstruct the data to a reasonable approximation. Currently investigators are required to search through the principal components in order to manually discover which ones contain features of the sample. This introduces subjectivity in to the analysis. Finally, as the data is transformed into linear combinations of the original data, interpretation of the components is not straight forward [22] (owing to abstract orthogonal, rather than meaningful chemical, factors), and particularly challenging for higher order components [23]. Specifically for MSI, negative scores are difficult to relate to mass spectral intensity [18], [24].

Others algorithms such as non-negative matrix factorization (NMF) are also used for dimensionality reduction and feature extraction of MSI data [25]. Although successful, like PCA, this method is linear, and the selection of the number of features k, is an unknown parameter. Random projection is a fast algorithm that can also reduce dimensionality [26], though results in the loss of information. Clustering techniques are used for spatial segmentation, however methods such as hierarchical clustering are limited to datasets with a small number of spectra as they require an p by p distanced matrix, where p is the number of pixels, to be stored in memory. Some efforts have been made to develop memory efficient algorithms in order to deal with the computationally prohibitive data size [10], though more effort is required.

Recently, t-distributed stochastic neighbour embedding (tSNE) [27] has been used for non-linear dimensional reduc-

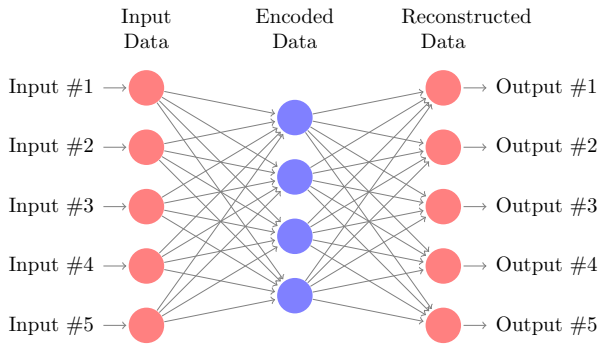


Fig. 2. The autoencoder consisting of an input layer, an encoding layer and an output layer where the input data is reconstructed from the encoded layer.

tion of MSI data [28]. Although in the original implementation the complexity scales with the number of pixels, p, squared, and becomes computationally prohibitive for large MS images. A more recent implementation uses the Barnes-Hut approximation and can scale with $p \log p$ after appropriately tuning an algorithmic parameter [29]. Currently, however, this method requires an initial dimensionality reduction of the data using PCA to alleviate the high computational costs.

Support vector machines (SVM) [30], genetic algorithms in combination with k-nearest neighbours [31], and SVM with neural networks [32] have also been used to analyse MSI data, though these have been used in only a few studies for the classification problems.

III. Method

The autoencoder is a class of neural network that are structurally symmetric, see Fig. 2, with the same number of nodes in the input and output layers. The input data is first encoded into the hidden layer, then reconstructed in the output layer. Typically, neural networks match an input to an classification and require explicit labelling of the data, i.e. supervised training, to assess the performance. As the aim in the autoencoder is to reconstruct the input data we have these labels implicitly. Therefore the necessary encoding and decoding transformations can be learned in an unsupervised manner.

The input vector $\mathbf{x}$ is encoded to the latent variable space $\mathbf{z}$ by applying a weight matrix $\mathbf{W}$ and bias vector $\mathbf{b}$ followed by applying an activation function, σ,

$$\mathbf{z} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}) \ . \tag{1}$$

Typically σ is a sigmoid function, hyperbolic tangent or linear transform. Here we use a sigmoid activation function in all cases. By encoding the input data into a reduced number of neurons it is possible to identify latent variables, i.e. generalised features, in the data characterised by $\mathbf{z}$.

The latent variables $\mathbf{z}$ are then decoded by applying another weight matrix ($\mathbf{W}'$), bias vector ($\mathbf{b}'$) and activation function, σ',

$$\mathbf{x}' = \sigma'(\mathbf{W}'\mathbf{z} + \mathbf{b}') \ . \tag{2}$$

Beginning with random elements in $\mathbf{W}$, $\mathbf{b}$, $\mathbf{W}'$ and $\mathbf{b}'$ we can define an unsupervised optimisation problem. The goal is to determine the elements in $\mathbf{W}$, $\mathbf{b}$, $\mathbf{W}'$ and $\mathbf{b}'$ that minimise the difference between $\mathbf{x}$ and $\mathbf{x}'$. Typically this reconstruction error (ϵ) is defined as the squared difference between $\mathbf{x}$ and $\mathbf{x}'$

$$\epsilon = ||\mathbf{x} - \mathbf{x}'||^2 = ||\mathbf{x} - \sigma'(\mathbf{W}'(\sigma(\mathbf{W}\mathbf{x} + \mathbf{b})) + \mathbf{b}')||^2 \ . \tag{3}$$

The elements of $\mathbf{W}$, $\mathbf{b}$, $\mathbf{W}'$ and $\mathbf{b}'$ can be iteratively updated via gradient decent and back propagation.

A. Regularised Sparse Autoencoders

In order to avoid over-fitting the training data we employ a standard L2 norm weight regularisation for w_{ij}, the elements of $\mathbf{W}$. This is calculated for L hidden nodes, n training examples and k variables in the training data, here the spectral channels.

$$\Omega_w = \frac{1}{2} \sum_{l}^{L} \sum_{j}^{n} \sum_{i}^{k} \left(w_{ji}^{(l)} \right)^2 \tag{4}$$

Additionally we include sparsity regularisation using the Kullback-Leibler (KL) divergence based on a desired level of sparsity ρ, given as

$$\Omega_s = \sum_{i=1}^{D'} \rho \log \left(\frac{\rho}{\hat{\rho}_i} \right) + (1 - \rho) \log \left(\frac{1 - \rho}{1 - \hat{\rho}_i} \right) \ , \tag{5}$$

where D' is the number of reduced dimensions and ρ is set to 0.5. Putting Eq. (3), with Eq. (4) and Eq. (5) yields a cost function of

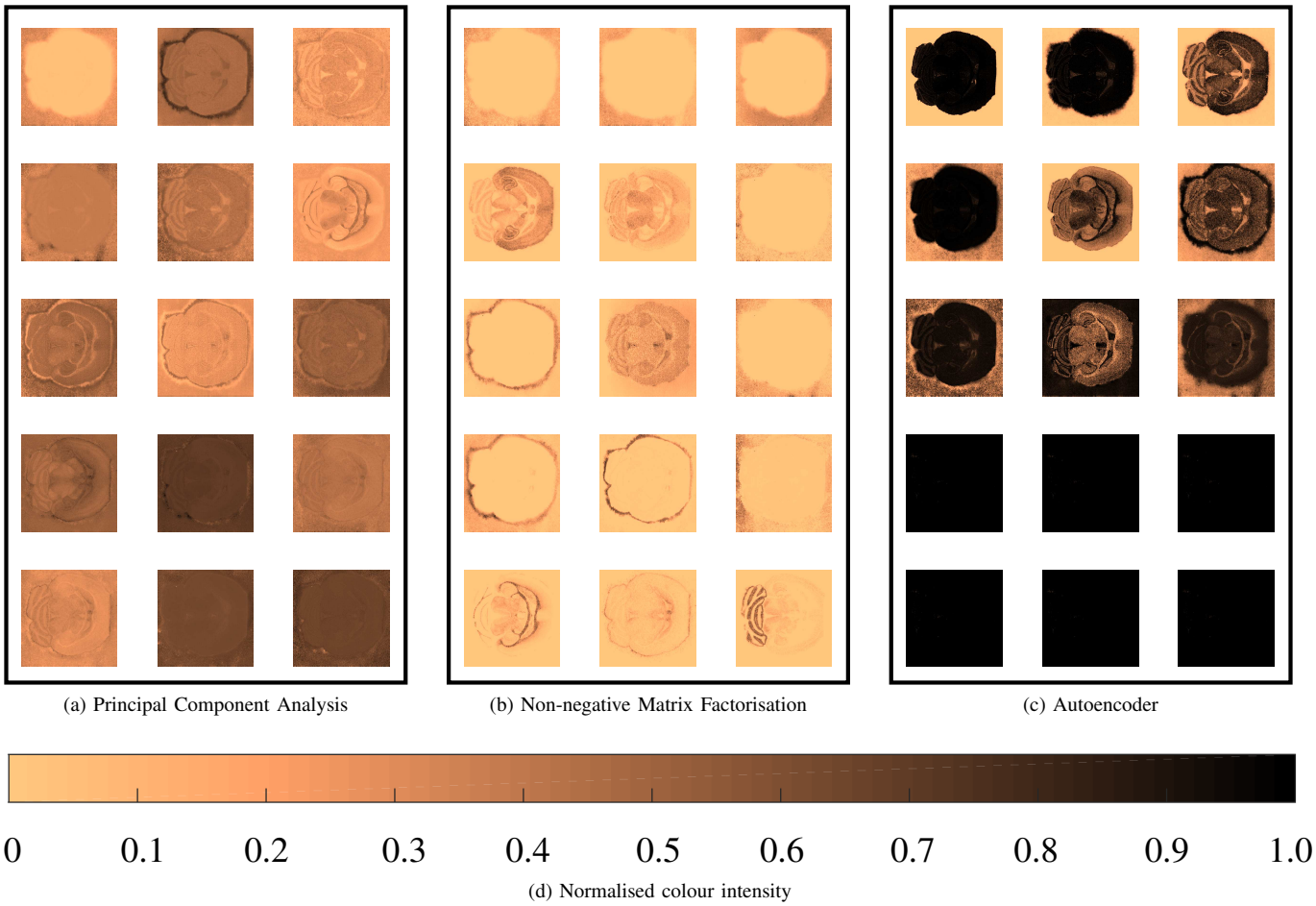
$$E = \frac{1}{n} \sum_{j=1}^{n} \epsilon_j + \lambda \Omega_w + \beta \Omega_s \tag{6}$$

to be minimised. Here ϵ_j is the outcome of Eq. (3) for the jth training example. We set the weight (λ) and sparsity (β) penalties to be 0.01 and 0.5 respectively.

B. Related Work

Autoencoders have been used in medical studies to classify data [33], [34], and extract features from times series data [35]. In these investigations, the training set consisted of a number of images where each image pixel is a single intensity value as opposed to hyper-spectral data. A very recent survey by Galli et al [25] reviewed the use of machine learning techniques in MALDI MSI data, though does not include autoencoders.

Recently studies have used autoencoders to extract features from hyper-spectral data [36], [37]. However, these data contained a comparable number of pixels, but only 100-200 spectral values, compared to the 1000s in MSI data. Furthermore, they use PCA to reduce the data to the first few principal components prior to using the autoencoder to learn spectral features in the data. However, as discussed in Section II-A, for MSI data the first principal component often



(a) Principal Component Analysis (b) Non-negative Matrix Factorisation (c) Autoencoder

(d) Normalised colour intensity

Fig. 3. (a) Principal component analysis (PCA), (b) non-negative matrix factorisation (NMF) and (c) Latent variables from the autoencoder of the mouse brain dataset. (d) Normalised colour intensities in range from 0 (copper) to 1 (black) for all methods for clarity of visualisation. Note features in each case have been ranked in order of decreasing variance.

separates the sample from the substrate so does not necessarily contain information about the sample features unlike other hyper-spectral data.

In this work we consider one 409x404 pixel image where each pixel is a mass spectrum containing 7036 *m/z* values. The spectral domain in the studies listed above are less than 2% of this size, with a comparable number of pixels. As autoencoders do not discard information during compression [38] it is preferential to use this technique without a prior reduction to avoid the loss of information that may be presents in high order principal components.

We train our autoencoder on the spectral data in order to identify the features that correspond to the regions in the ion images. That is we encode the 7036 intensities that make up a mass spectrum for each pixel, for which we have 165,236 (409x404) training examples. Our autoencoder consists of 7036 input nodes and 15 hidden nodes, taking each of the 165,236 pixels as a training example. We train the autoencoder only in an unsupervised manor and do not use supervised fine tuning.

IV. Data: Mouse Brain

We apply our autoencoder to a MALDI MSI dataset of a transverse section of a mouse brain. The data were acquired from a Waters Synapt G2-Si and pre-processed using SpectralAnalysis [39]. The raw data are processed such that the spectral channels undergo zero filling based on the set of union of all *m/z* bins to create a consistent axis, followed by Savitzky-Golay smoothing to reduce noise [10]. The spectra are then 'peak picked' based on a noise threshold. After pre-processing the dataset contains 409X404 pixels each with mass spectrum of 7036 *m/z* intensity values yielding a dataset of 8GB.

It is common to normalise or scale the data, though various techniques exist and there is no consensus in the field. The resulting analysis depends on the differing normalisation techniques, which can significantly affect the outcome [24], [40], [41]. A standard method of normalisation is based on the total ion count in the spectra and is used here prior to any dimensional reduction.

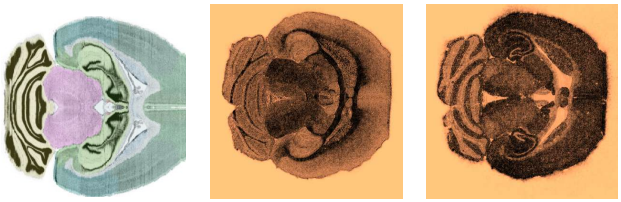
V. Results

The ion images here are viewed in a colour scheme in order to avoid over or under emphasising features in the data following the recommendations in Race *et al.* [42]. We compare the results of our autoencoder to that of principal component analysis (PCA) and non-negative matrix factorisation (NMF). In order directly compare the methods, we normalise the reduced dimensional data to the most intense feature in each pixel. That is, the colour (intensity) of the lower dimensional images range from 0 to 1 for all pixels in all methods.

The scores for PCA are given in Fig. 3(a) for the first 15 components, which explain 99.37% of the variance in the data. However, the first 5 components account for 82.28%, 12.83%, 3.21%, 0.62% and 0.44% of the variance respectively. Note that the results of PCA produce positive and negative scores. As the negative scores are difficult to interpret, we show only the positive scores for PCA. Figure 3(b) shows the results for the NMF algorithm with 15 factors. NMF is able to extract separate the features in the dataset and has the additional benefit of the corresponding spectra for each feature. These spectra contain peaks of varying intensities so yields normalisation issues as mentioned in Section IV. If one normalises to the most intense peak to provide a comparable range with other methods, the over all ion image becomes very bight as it is dominated by a few pixels. As a consequence the analysis of these images is not straight forward and may be difficult to interpret in general.

The results of the dimensional reduction using the autoencoder described in Section III are given in Fig. 3(c). Similarly to PCA and NMF, the autoencoder generates several uninformative features in the data. Recall Section IV where many peaks exist in the spectral dimension, though only a few of them correspond to the sample. Many of the other spectral peaks are related to the substrate or to the matrix used in the MALDI sample preparation. As a result, many of the features in the data are due to 'noise' and are uninformative. Even the peaks corresponding to the matrix can be considered noise in this context as the signal is approximately uniform over all pixels (this is a direct result of the matrix being sprayed evenly over the sample in the experimental preparation). This also explains why many of the features in PCA are also trivial in the context of the sample. The frequency of trivial features may be due to the high levels of redundancy in spectral information [37], with multiple sources of noise leading to several 'noise' features. We note the autoencoder generates uniformity image intensities for these trivial features thus automated detected and removal of these features for larger studies would be straightforward.

It is worth noting that the autoencoder has compressed much of the spectral information in to a few informative features. That is, the hidden neurons that contain information relating to the sample capture almost all of the regions of the brain, see Fig. 4. This is in contrast to PCA and NMF, which tend to separate regions into different features, where reduced dimensions need to be combined to achieve a single representation of the data. In terms of performance on this problem, the autoencoder has reduced the MSI data into a few components that capture the structural information of the sample, and several components that relate to noise (either substrate or matrix). Figure 4 illustrates the ability of the autoencoder to separate the anatomical regions of the brain based on unsupervised dimensionality reduction where most regions are captured by a single hidden node. This demonstrates the success of the autoencoder at reducing the dimensionality of MSI data to a few *m/z* values that provide structural information about the sample.



(a) Brain Anatomy (b) AE latent variable (c) AE latent variable

Fig. 4. Comparison of (a) known regions of a mouse brain from the Allen Brain Atlas [43], [44] to (b) and (c) a single hidden neurons in the autoencoder.

VI. Conclusions

Here we have applied the autoencoder to MSI data for unsupervised non-linear dimensionality reduction. Only a few investigations relate to the use of autoencoders with multi-dimensional data. These either consist of times series datasets that are treated as more training examples, rather than additional dimension in the data, or a small number of spectral channels that are initially reduced via PCA. We successfully apply an autoencoder to the mass spectrum for each pixel in an MS image and reduce it to its core features; structural information about the sample and noise. In contrast to the standard PCA and NMF techniques, the autoencoder provides a method for non-linear dimensionality reduction. Moreover, as each pixel (containing a mass spectrum) is treated as a training example, the whole dataset does not need to be loaded in memory at once. Also, provided that the number of hidden neurons is less that the number of channels in the mass spectrum, the autoencoder will be less computationally expensive than t-SNE or clustering methods. This may restrict the use of the so called over-complete representation of the data in stochastic denoising autoencoders which can learn 'deeper' features in the data [45]. However, the 'over-complete' hidden layer, which transforms the data it into a sparse representation, is unnecessary for MSI as the spectral data are inherently sparse.

Although training the weight matrix requires a large amount of computation time (around 20hrs for the autoencoder compared to 1hr for PCA and 1.5hrs for NMF), once obtained the encoding and decoding of the data is fast (10s of seconds). Therefore, when considering multiple samples e.g. control studies, analysing databases, and replicates though repeated sample acquisition [46] or serial sections, the autoencoder

will be much more time efficient than other dimensionality reduction techniques which would need to be applied to individual datasets. Moreover, analysing data separately with existing methods introduces difficulties in comparison between datasets due variations in the transform from the high dimensional to low dimensional space. With a fixed mapping to and from the reduced dimensional space, it is possible to use the autoencoder to encode multiple datasets to the same number of reduced dimensions, and through the same transformation, enabling direct comparison between datasets.

The number of hidden nodes in the autoencoder could be viewed as a tunable parameter similar to the number of features determined through NMF. However, as demonstrated here, the autoencoder combines informative features to a few hidden nodes, rather than each detecting a different features as in PCA and NMF. As a result it is not necessary to tune the number of hidden nodes.

As the application of autoencoders to hyper-spectral data has been largely unexplored, investigations into algorithmic performance is an obvious starting point. Implementation of training mini-batches [47] could significantly improve performance when computed in parallel over multiple cores. However one must be able to store the required data (e.g. the weight matrix) in memory, or develop memory-efficient implementations for this to be feasible. More investigation into the interpretation of the encoded spectral information, and how it relates to the sample, is essential for uptake in the MS community.

ACKNOWLEDGMENT

This work forms part of the Strategic Capability Programme (projects "3D nanoSIMS" and "AIMS-HIGHER") of the National Measurement System of the UK Department of Business, Energy and Industrial Strategy.

REFERENCES

[1] M. Karas, D. Bachmann, and F. Hillenkamp, "Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules," *Analytical Chemistry*, vol. 57, no. 14, pp. 2935–2939, 1985.

[2] Z. Takáts, J. M. Wiseman, B. Gologan, and R. G. Cooks, "Mass spectrometry sampling under ambient conditions with desorption electrospray ionization," *Science*, vol. 306, no. 5695, pp. 471–473, 2004.

[3] C. W. Magee, W. L. Harrington, and R. E. Honig, "Secondary ion quadrupole mass spectrometer for depth profilingdesign and performance evaluation," *Review of Scientific Instruments*, vol. 49, no. 4, 1978.

[4] M. Stoeckli, D. Staab, and A. Schweitzer, "Compound and metabolite distribution measured by MALDI mass spectrometric imaging in whole-body tissue sections," *International Journal of Mass Spectrometry*, vol. 260, no. 23, pp. 195 – 202, 2007, imaging Mass Spectrometry Special Issue.

[5] D. S. Cornett, S. L. Frappier, and R. M. Caprioli, "MALDI-FTICR imaging mass spectrometry of drugs and metabolites in tissue," *Analytical Chemistry*, vol. 80, no. 14, pp. 5648–5653, 2008.

[6] C. L. Carter, C. W. McLeod, and J. Bunch, "Imaging of phospholipids in formalin fixed rat brain sections by matrix assisted laser desorption/ionization mass spectrometry," *Journal of The American Society for Mass Spectrometry*, vol. 22, no. 11, pp. 1991–1998, 2011.

[7] A. D. Palmer, R. Griffiths, I. Styles, E. Claridge, A. Calcagni, and J. Bunch, "Sucrose cryo-protection facilitates imaging of whole eye sections by MALDI mass spectrometry," *Journal of Mass Spectrometry*, vol. 47, no. 2, pp. 237–241, 2012.

[8] I. M. Taban, A. F. M. Altelaar, Y. E. M. van der Burgt, L. A. McDonnell, R. M. A. Heeren, J. Fuchser, and G. Baykut, "Imaging of peptides in the rat brain using MALDI-FTICR mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 18, no. 1, pp. 145–151, 2007.

[9] S. Khatib-Shahidi, M. Andersson, J. L. Herman, T. A. Gillespie, and R. M. Caprioli, "Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry," *Analytical Chemistry*, vol. 78, no. 18, pp. 6448–6456, 2006.

[10] A. M. Race, R. T. Steven, A. D. Palmer, I. B. Styles, and J. Bunch, "Memory efficient principal component analysis for the dimensionality reduction of large mass spectrometry imaging data sets," *Analytical Chemistry*, vol. 85, no. 6, pp. 3071–3078, 2013.

[11] A. Römpp, R. Wang, J. P. Albar, A. Urbani, H. Hermjakob, B. Spengler, and J. A. Vizcano, "A public repository for mass spectrometry imaging data," *Analytical and bioanalytical chemistry*, vol. 407, no. 8, p. 20272033, March 2015.

[12] R. J. Carreira, R. Shyti, B. Balluff, W. M. Abdelmoula, S. H. van Heiningen, R. J. van Zeijl, J. Dijkstra, M. D. Ferrari, E. A. Tolner, L. A. McDonnell, and A. M. J. M. van den Maagdenberg, "Large-scale mass spectrometry imaging investigation of consequences of cortical spreading depression in a transgenic mouse model of migraine," *Journal of The American Society for Mass Spectrometry*, vol. 26, no. 6, pp. 853–861, 2015.

[13] L. A. Klerk, A. Broersen, I. W. Fletcher, R. van Liere, and R. M. Heeren, "Extended data analysis strategies for high resolution imaging ms: New methods to deal with extremely large image hyperspectral datasets," *International Journal of Mass Spectrometry*, vol. 260, no. 23, pp. 222 – 236, 2007, imaging Mass Spectrometry Special Issue.

[14] P. Källback, A. Nilsson, M. Shariatgorji, and P. E. Andrén, "msiquant quantitation software for mass spectrometry imaging enabling fast access, visualization, and analysis of large data sets," *Analytical Chemistry*, vol. 88, no. 8, pp. 4346–4353, 2016.

[15] J. C. May and J. A. McLean, "Advanced multidimensional separations in mass spectrometry: Navigating the big data deluge," *Annual Review of Analytical Chemistry*, vol. 9, no. 1, pp. 387–409, 2016.

[16] P. J. Cumpson, N. Sano, I. W. Fletcher, J. F. Portoles, M. Bravo-Sanchez, and A. J. Barlow, "Multivariate analysis of extremely large tofsims imaging datasets by a rapid pca method," *Surface and Interface Analysis*, vol. 47, no. 10, pp. 986–993, 2015.

[17] J. Yang, O. Rbel, Prabhat, M. W. Mahoney, and B. P. Bowen, "Identifying important ions and positions in mass spectrometry imaging data using cur matrix decompositions," *Analytical Chemistry*, vol. 87, no. 9, pp. 4658–4666, 2015.

[18] T. Alexandrov, "MALDI imaging mass spectrometry: statistical data analysis and current computational challenges," *BMC Bioinformatics*, vol. 13, no. 16, pp. 1–13, 2012.

[19] T. Alexandrov, M. Becker, S.-O. Deininger, G. Ernst, L. Wehder, M. Grasmair, F. von Eggeling, H. Thiele, and P. Maass, "Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering," *Journal of Proteome Research*, vol. 9, no. 12, pp. 6535–6546, 2010.

[20] T. Hayasaka, N. Goto-Inoue, M. Ushijima, I. Yao, A. Yuba-Kubo, M. Wakui, S. Kajihara, M. Matsuura, and M. Setou, "Development of imaging mass spectrometry (IMS) dataset extractor software, IMS convolution," *Analytical and Bioanalytical Chemistry*, vol. 401, no. 1, pp. 183–193, 2011.

[21] A. G. Shard, S. J. Spencer, S. A. Smith, R. Havelund, and I. S. Gilmore, "The matrix effect in organic secondary ion mass spectrometry," *International Journal of Mass Spectrometry*, vol. 377, pp. 599 – 609, 2015, special Issue: MS 1960 to Now.

[22] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, vol. 12, pp. 265–286, 2006.

[23] G. I. Allen, "Sparse higher-order principal components analysis," in *In Proceedings of 15th International Conference on Artificial Intelligence and Statistics*, 2012, pp. 27–36.

[24] J. L. S. Lee, I. S. Gilmore, I. W. Fletcher, and M. P. Seah, "Multivariate image analysis strategies for tof-sims images with topography," *Surface and Interface Analysis*, vol. 41, no. 8, pp. 653–665, 2009.

[25] M. Galli, I. Zoppis, A. Smith, F. Magni, and G. Mauri, "Machine learning approaches in MALDI-MSI: clinical applications," *Expert Review of Proteomics*, vol. 13, no. 7, pp. 685–696, 2016.

[26] A. D. Palmer, J. Bunch, and I. B. Styles, "The use of random projections for the analysis of mass spectrometry imaging data," *Journal of The American Society for Mass Spectrometry*, vol. 26, no. 2, pp. 315–322, 2015.

[27] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[28] W. M. Abdelmoula, K. krkov, B. Balluff, R. J. Carreira, E. A. Tolner, B. P. F. Lelieveldt, L. van der Maaten, H. Morreau, A. M. J. M. van den Maagdenberg, R. M. A. Heeren, L. A. McDonnell, and J. Dijkstra, "Automatic generic registration of mass spectrometry imaging data to histology using nonlinear stochastic embedding," *Analytical Chemistry*, vol. 86, no. 18, pp. 9204–9211, 2014.

[29] L. van der Maaten, "Barnes-Hut-SNE," *CoRR*, vol. abs/1301.3342, 2013. [Online]. Available: http://arxiv.org/abs/1301.3342

[30] M. R. Groseclose, P. P. Massion, P. Chaurand, and R. M. Caprioli, "High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry," *PROTEOMICS*, vol. 8, no. 18, pp. 3715–3724, 2008.

[31] L. H. Cazares, D. Troyer, S. Mendrinos, R. A. Lance, J. O. Nyalwidhe, H. A. Beydoun, M. A. Clements, R. R. Drake, and O. J. Semmes, "Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase kinase 2 discriminates cancer from uninvolved prostate tissue," *American Association for Cancer Research*, vol. 15, no. 17, pp. 5541–5551, 2009.

[32] S. Rauser, C. Marquardt, B. Balluff, S.-O. Deininger, C. Albers, E. Belau, R. Hartmer, D. Suckau, K. Specht, M. P. Ebert, M. Schmitt, M. Aubele, H. Höfler, and A. Walch, "Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry," *Journal of Proteome Research*, vol. 9, no. 4, pp. 1854–1863, 2010.

[33] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans," *Sci Rep*, pp. 1–13, 2016.

[34] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 1077–1089, April 2016.

[35] H.-C. Shin, M. Orton, D. J. Collins, S. Doran, and M. O. Leach, "Autoencoder in time-series analysis for unsupervised tissues characterisation in a large unlabelled medical image dataset," in *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, vol. 1, 2011, pp. 259–264.

[36] C. Tao, H. Pan, Y. Li, and Z. Zou, "Unsupervised spectral-spatial feature learning with stacked sparse autoencoders for hyperspectral imagery classification," *IEEE Geoscience and remote sensing letters*, vol. 12, no. 12, pp. 2438–2442, 2015.

[37] L. Wang, J. Zhang, P. Liu, K.-K. R. Choo, and F. Huang, "Spectral–spatial multi-feature-based deep learning for hyperspectral remote sensing image classification," *Soft Computing*, pp. 1–9, 2016.

[38] S. J. Hong, G. S. May, and D.-C. Park, "Neural network modeling of reactive ion etching using optical emission spectroscopy data," *IEEE Transactions on Semiconductor Manufacturing*, vol. 16, no. 4, pp. 598–608, Nov 2003.

[39] A. M. Race, A. D. Palmer, A. Dexter, R. T. Steven, I. B. Styles, and J. Bunch, "Spectralanalysis: software for the masses," *Analytical Chemistry*, vol. 88, pp. 9541–9458, 2016.

[40] J. M. Fonville, C. Carter, O. Cloarec, J. K. Nicholson, J. C. Lindon, J. Bunch, and E. Holmes, "Robust data processing and normalization strategy for MALDI mass spectrometric imaging," *Analytical Chemistry*, vol. 84, no. 3, pp. 1310–1319, 2012.

[41] J. L. S. Lee, I. S. Gilmore, and M. P. Seah, "Quantification and methodology issues in multivariate analysis of tof-sims data for mixed organic systems," *Surface and Interface Analysis*, vol. 40, no. 1, pp. 1–14, 2008.

[42] A. M. Race and J. Bunch, "Optimisation of colour schemes to accurately display mass spectrometry imaging data based on human colour perception," *Analytical and Bioanalytical Chemistry*, vol. 407, no. 8, pp. 2047–2054, 2015.

[43] E. S. Lein, M. J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A. F. Boe, M. S. Boguski, K. S. Brockway, E. J. Byrnes, L. Chen, L. Chen, T.-M. Chen, M. Chi Chin, J. Chong, B. E. Crook, A. Czaplinska, C. N. Dang, S. Datta, N. R. Dee, A. L. Desaki, T. Desta, E. Diep, T. A. Dolbeare, M. J. Donelan, H.-W. Dong, J. G. Dougherty, B. J. Duncan, A. J. Ebbert, G. Eichele, L. K. Estin, C. Faber, B. A. Facer, R. Fields, S. R. Fischer, T. P. Fliss, C. Frensley, S. N. Gates, K. J. Glattfelder, K. R. Halverson, M. R. Hart, J. G. Hohmann, M. P. Howell, D. P. Jeung, R. A. Johnson, P. T. Karr, R. Kawal, J. M. Kidney, R. H. Knapik, C. L. Kuan, J. H. Lake, A. R. Laramee, K. D. Larsen, C. Lau, T. A. Lemon, A. J. Liang, Y. Liu, L. T. Luong, J. Michaels, J. J. Morgan, R. J. Morgan, M. T. Mortrud, N. F. Mosqueda, L. L. Ng, R. Ng, G. J. Orta, C. C. Overly, T. H. Pak, S. E. Parry, S. D. Pathak, O. C. Pearson, R. B. Puchalski, Z. L. Riley, H. R. Rockett, S. A. Rowland, J. J. Royall, M. J. Ruiz, N. R. Sarno, K. Schaffnit, N. V. Shapovalova, T. Sivisay, C. R. Slaughterbeck, S. C. Smith, K. A. Smith, B. I. Smith, A. J. Sodt, N. N. Stewart, K.-R. Stumpf, S. M. Sunkin, M. Sutram, A. Tam, C. D. Teemer, C. Thaller, C. L. Thompson, L. R. Varnam, A. Visel, R. M. Whitlock, P. E. Wohnoutka, C. K. Wolkey, V. Y. Wong, M. Wood, M. B. Yaylaoglu, R. C. Young, B. L. Youngstrom, X. Feng Yuan, B. Zhang, T. A. Zwingman, and A. R. Jones, "Genome-wide atlas of gene expression in the adult mouse brain," *Nature*, vol. 445, pp. 168–176, 2007.

[44] A. B. Atlas, "Anatomic Gene Expression Atlas (AGEA) ," http://mouse.brain-map.org/agea, [Online; accessed 11-August-2016].

[45] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 1096–1103.

[46] R. T. Steven and J. Bunch, "Repeat MALDI MS imaging of a single tissue section using multiple matrices and tissue washes," *Analytical and Bioanalytical Chemistry*, vol. 405, no. 14, pp. 4719–4728, 2013.

[47] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.