# Deep variational auto-encoders for unsupervised glomerular classification

Brendon Lutnick, Rabi Yacoub, Kuang-Yu Jen, John Tomaszewski, Sanjay Jain, et al.

Brendon Lutnick, Rabi Yacoub, Kuang-Yu Jen, John E. Tomaszewski, Sanjay Jain, Pinaki Sarder, "Deep variational auto-encoders for unsupervised glomerular classification," Proc. SPIE 10581, Medical Imaging 2018: Digital Pathology, 105810C (6 March 2018); doi: 10.1117/12.2295456

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

# Deep variational auto-encoders for unsupervised glomerular classification

Brendon Lutnick[1], Rabi Yacoub[2], Kuang-Yu Jen[3], John E. Tomaszewski[1], Sanjay Jain[4], and Pinaki Sarder[1,*]

[1]Department of Pathology and Anatomical Sciences,
[2]Medicine – Nephrology,
University at Buffalo – The State University of New York

[3]Department of Pathology,
University of California at Davis

[4]Department of Medicine – Nephrology,
Washington University School of Medicine

*Address all correspondence to: Pinaki Sarder, Tel: 716-829-2265, email: pinakisa@buffalo.edu

## ABSTRACT

The adoption of deep learning techniques in medical applications has thus far been limited by the availability of the large labeled datasets required to robustly train neural networks, as well as difficulty interpreting these networks. However, recent techniques for unsupervised training of neural networks promise to address these issues, leveraging only structure to model input data. We propose the use of a variational autoencoder (VAE) which utilizes data from an animal model to augment the training set and non-linear dimensionality reduction to map this data to human sets. This architecture utilizes variational inference, performed on latent parameters, to statistically model the probability distribution of training data in a latent feature space. We show the feasibility of VAEs, using images of mouse and human renal glomeruli from various pathological stages of diabetic nephropathy (DN), to model the progression of structural changes which occur in DN. When plotted in a 2-dimentional latent space, human and mouse glomeruli, show separation with some overlap, suggesting that the data is continuous, and can be statistically correlated. When DN stage is plotted in this latent space, trends in disease pathology are visualized.

**Keywords:** Cross species modeling, knowledge transfer, variational autoencoder, unsupervised modeling

## 1. INTRODUCTION

Clinical pathology relies on manual qualification of biological structure from images, which is time consuming and often error prone. Computer vision algorithms, which aid histopathological image analysis, will likely provide useful quantitative analysis, and serve as important pre-screening tools in the near future. However, large training sets are hard to come by, and expert annotation is costly. Algorithms, finely tuned to biological structure, will facilitate the transition to quantitatively informed analysis in clinical pathology, which has historically relied on manual qualification of biological structure. Given the importance of imaging in the field, these algorithms must be capable of performing highly non-linear dimensionality reduction prior to quantification of tissue slides. Traditionally data features are hand engineered, and while such practice works well for some datasets, automated schemes for dimensionality reduction have the potential to explore more intricate dimensional relationships than hand-selected features ever feasibly could. Currently, many consider deep learning the state-of-the-art machine learning technique; however, it requires exhaustively labeled training data sets and is often criticized for its black-box nature.

Despite the difficulty interpreting neural networks, the precise and accurate quantitative analysis provided by these machine learning techniques is increasingly seeing use in modern medicine and science, promising to discover important information about biological systems using microscopic imaging data. However, large annotated biological training sets are hard to come by, as expert annotation is costly. Animal model systems, commonly used for cost effective biological data generation, have limited direct translatability to human systems. There is a need for a robust quantitative method, able to map features across species, trainable with large unlabeled datasets. Towards this direction, we are developing a probabilistic unsupervised framework for generative data modeling which leverages dimensionality reduction to learn complex representations of the input data. We expect this method to provide unique insight into the relationship between animal models, and human data.

Medical Imaging 2018: Digital Pathology, edited by John E. Tomaszewski, Metin N. Gurcan,
Proc. of SPIE Vol. 10581, 105810C · © 2018 SPIE · CCC code:
1605-7422/18/$18 · doi: 10.1117/12.2295456

Of interest is modeling of structural features present in stained human kidney tissue. Particularly we aim to focus on the glomerulus, the primary waste filtering unit of the kidney, commonly examined for structural damage via renal biopsy. The glomerulus is a ball of entwined capillaries encapsulated by the Bowman's capsule. This structure is able to selectively control the permeability and specificity of filtration of blood plasma into Bowman's space, by various biological pathways[1]. We intend to use the expansive mouse model data acquired by our group to inform models of human glomeruli, where data is limited, using this data to augment human training sets to build automatic models of glomerular structural changes observed in diabetic nephropathy (DN). We hypothesize that non-linear latent representations human and murine biopsy imaging data will allow linear mapping between the sets, enabling accurate unsupervised modeling of these structural changes.

## 2. METHODS

We use a variational autoencoder architecture (VAE)[2, 3], trained using raw data from mouse and human renal tissue, which was imaged using common histopathological stains. By balancing a negative log-likelihood reconstruction cost with the Kullback-Leibler divergence[4] between the latent variables and a known distribution, this network strikes a balance between compression efficiency and data fidelity. Of interest, is the latent (central) layer in the network where images are encoded to a highly reduced probabilistic space. VAE architectures are probabilistic models, which learn the parameters of the statistical (Gaussian) distributions, which define their latent layers, keeping the network differentiable and therefore trainable by stochastic gradient decent[5].

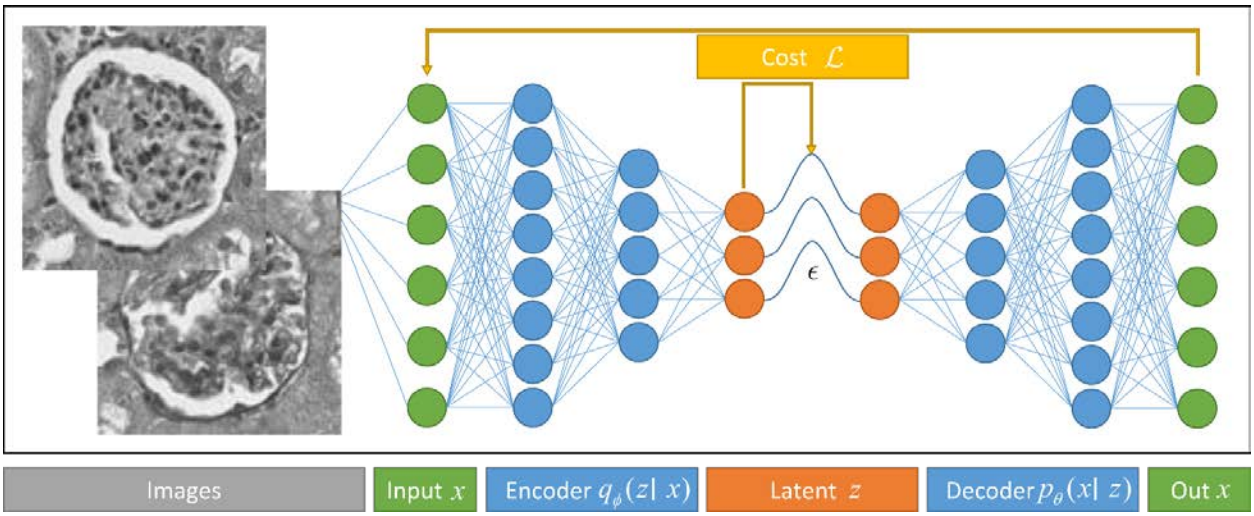


**Figure 1. Example of a fully connected VAE and the glomerular training images used.** The encoder network is coupled to a mirrored decoder by statistical sampling performed on the latent distributions (Gaussian functions), whose parameters are learned using variational inference. The network is penalized for poor reconstruction, and regularized using the KL divergence between the latent distributions and a prior (unit Gaussian).

### 2.1. VAE cost function derivation

The unsupervised generative modeling by VAE architectures is a result of a decoupling of the encoder and decoder, through latent sampling, and can be explained through a Bayesian lens.

Let us define $\boldsymbol{D}$ as the data matrix or grayscale image with dimension $M \times N$, where $(m, n)$ denotes a pixel location. The network is trained using $K$ images $\{\boldsymbol{D}_1, \boldsymbol{D}_2, \dots, \boldsymbol{D}_K\}$, and we denote $\boldsymbol{x}_k = vec(\boldsymbol{D}_k)$. In the following, we drop the subscripts and use $\boldsymbol{x}$ to denote any one of the $K$ input images.

We use the following notations to derive the network cost: $\boldsymbol{I}$ is the identity matrix, $\odot$ is the Hadamard product, $E$ is the expectation operation, and $p(\cdot)$ denotes a probability density function.

Unlike traditional autoencoder architectures, which learn direct latent representations $\boldsymbol{z}$ of the data $\boldsymbol{x}$, VAEs learn the approximate posterior $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})$ from which the latent variables are sampled. To ensure that the model is differentiable (trainable using gradient decent) with respect to the learned parameters, a probabilistic sampling node is included:

$$\boldsymbol{\epsilon} = \mathcal{N}(0, \boldsymbol{I}), \quad (1)$$

assuming:

$$q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}), \quad (2)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{I}$ have the same dimension as the latent space. The latent sampling is defined as:

$$\boldsymbol{z} = \boldsymbol{\mu} + \sigma \odot \boldsymbol{\epsilon}, \quad (3)$$

where the parameters $\boldsymbol{\mu}$ and $\sigma$ are predicted by the network. The latent variables are then decoded to give the output, $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})$. The network is optimized by maximizing the evidence lower bound (ELBO) $\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}; \boldsymbol{x})$:

$$\mathcal{L}(\cdot) = E_{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})] - \mathcal{D}_{KL}\big(q_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{x}) \parallel p(\boldsymbol{z})\big), \quad (4)$$

where, $E_{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})}[\cdot]$ represents the quality of reconstruction, and the Kullback-Leibler (KL) divergence[4] $\mathcal{D}_{KL}(\cdot)$ penalizes the difference between the approximate and true latent posterior distributions, serving to regularize the model. The model is optimized by maximizing ELBO and can be conceptualized heuristically as balancing the cost of poor reconstruction, with poor generalization.

This cost function is easily derived using Bayesian statistics, where the convenient prior: $\boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I})$ is enforced over the latent variables. Here the generative network is defined by:

$$p(\boldsymbol{x}, \boldsymbol{z}) = p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z}), \quad (5)$$

and the network inference as:

$$p(\boldsymbol{z}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}{p(\boldsymbol{x})}. \quad (6)$$

Here, the probability of the true data posterior distribution $p(\boldsymbol{x}) = \int p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})d\boldsymbol{x}$ is intractable, and therefore must be approximated as $\boldsymbol{\lambda}_{\boldsymbol{x}_i} = \big(\boldsymbol{\mu}_{\boldsymbol{x}_i}, \boldsymbol{\sigma}_{\boldsymbol{x}_i}\big)$. The KL divergence[4] between the inference network and encoder is used to determine the accuracy of the model:

$$\mathcal{D}_{KL}\big(q_{\boldsymbol{\lambda}}(\boldsymbol{z}|\boldsymbol{x}) \parallel p(\boldsymbol{z}|\boldsymbol{x})\big) = \int q_{\boldsymbol{\lambda}}(\boldsymbol{z}|\boldsymbol{x}) \log \frac{q_{\boldsymbol{\lambda}}(\boldsymbol{z}|\boldsymbol{x})}{p(\boldsymbol{z}|\boldsymbol{x})}, \quad (7)$$

which can be written as:

$$\mathcal{D}_{KL}(\cdot) = E_q[\log q_{\boldsymbol{\lambda}}(\boldsymbol{z}|\boldsymbol{x})] - E_q[\log p(\boldsymbol{x}, \boldsymbol{z})] + \log p(\boldsymbol{x}). \quad (8)$$

Here the goal is to find the variational parameters of $\boldsymbol{\lambda}$ which minimize $\mathcal{D}_{KL}(\cdot)$, ensuring the decoded data matches the original. Here we can define ELBO as:

$$\mathcal{L}(\boldsymbol{\lambda}) = E_q[\log p(\boldsymbol{x}, \boldsymbol{z})] - E_q[\log q_{\boldsymbol{\lambda}}(\boldsymbol{z}|\boldsymbol{x})], \quad (9)$$

so, the KL divergence becomes:

$$\log p(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{\lambda}) + \mathcal{D}_{KL}\big(q_{\boldsymbol{\lambda}}(\boldsymbol{z}|\boldsymbol{x}) \parallel p(\boldsymbol{z}|\boldsymbol{x})\big). \quad (10)$$

Using the property: $\mathcal{D}_{KL}(\cdot) \geq 0$, we can maximize ELBO to minimize $\mathcal{D}_{KL}(\cdot)$ without the need to compute $p(\boldsymbol{x})$. This is known as variational inference, where the variational parameters $\boldsymbol{\lambda}$ which define the posterior distribution $(\boldsymbol{z})$ are optimized. To implement this practically, the variational parameters are used to define $\boldsymbol{\mu}$ and $\sigma$ of the chosen prior distribution $\boldsymbol{z}$.

### 2.2. Network implementation

As proof of concept, we use histopathological glomerular images from mice and humans. For preliminary testing we use a fully connected network (Fig. 1) which has an architecture of 10000, 500, 500, 2, 500, 500, 10000 nodes, trained using a set of 14184, 100 × 100 vectorized black and white images, augmented by rotation and flipping to contain 113480. Here 2 latent nodes was chosen for ease of visualization, allowing for proof of concept. Practically we found that training with a batch size of 128 with a learning rate of 2e-5 for 50 epochs gave reproducible results.

### 2.3. Murine model

A standard streptozocin (STZ) treated mouse model[6] was used. C57BL/6J background mice (7 weeks old) were injected with STZ. They develop a mild form of diabetes mellitus type I, and after 25 weeks, mild-moderate diabetic nephropathy (DN), with untreated mice used as control. All animal studies were performed in accordance with protocols approved by the University at Buffalo Animal Studies Committee.

### 2.4. Human data

Biopsy samples from human diabetic DN patients with chronic kidney disease (CKD) stage II and stage III were collected from Kidney Translational Research Center at Washington University School of Medicine directed by co-author Dr. Sanjay Jain. The glomerular structural changes in these biopsies suggest DN related changes spanning different DN stages as discussed in Tervaert *et al.*[7]. As control, renal tissue samples of non-diabetic patients with no apparent histological abnormalities were considered. The image dataset was verified by Dr. Sanjay Jain and Dr. John E. Tomaszewski. Human data collection procedure followed a protocol approved by the Institutional Review Board at University at Buffalo. Ground-truth annotations of DN structural disease state was performed by the co-author Dr. Kuang-Yu Jen.

### 2.5. Imaging and data preparation

Tissue slices of 2 µm (for murine data) and 2-5 µm (for human data) were stained using Periodic acid-Schiff, and imaged using a whole-slide imaging scanner (Aperio Versa, Leica, Buffalo Grove, Il). We followed similar imaging protocol as described in our earlier works[8].

## 3. RESULTS

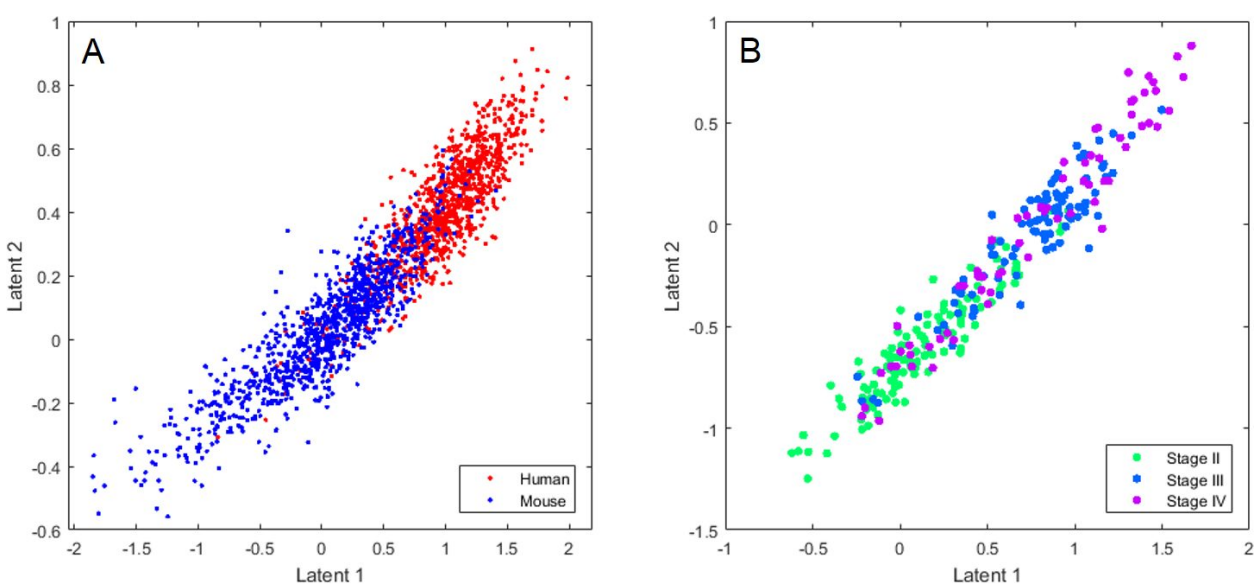


**Figure 2. Preliminary VAE human glomerular mapping results, plotting disease state, trained using unlabeled human and murine glomeruli.** (A) The distribution of mouse and human glomeruli images visualized in a 2-dimensional latent space by plotting labeled data. (B) Trends in human DN progression[7] are observed in data plotted in the same 2-dimentional latent space. We hypothesize that VAE mapping will provide fuzzy class labels reflecting a continuous spectrum of disease, rather than categorical classifications.

When testing using both the mouse and human data, clear separation between species is observed, with some overlap, suggesting cross-species knowledge transfer can be conducted between these two species (Fig. 2A). Additionally, despite limited human training data, a trend in diabetic nephropathy (DN) disease states from annotated human glomeruli can be observed (Fig. 2B). It is noteworthy to mention that these networks were trained using only unlabeled data, and the trends in training data sets are the result of the pure data structure.

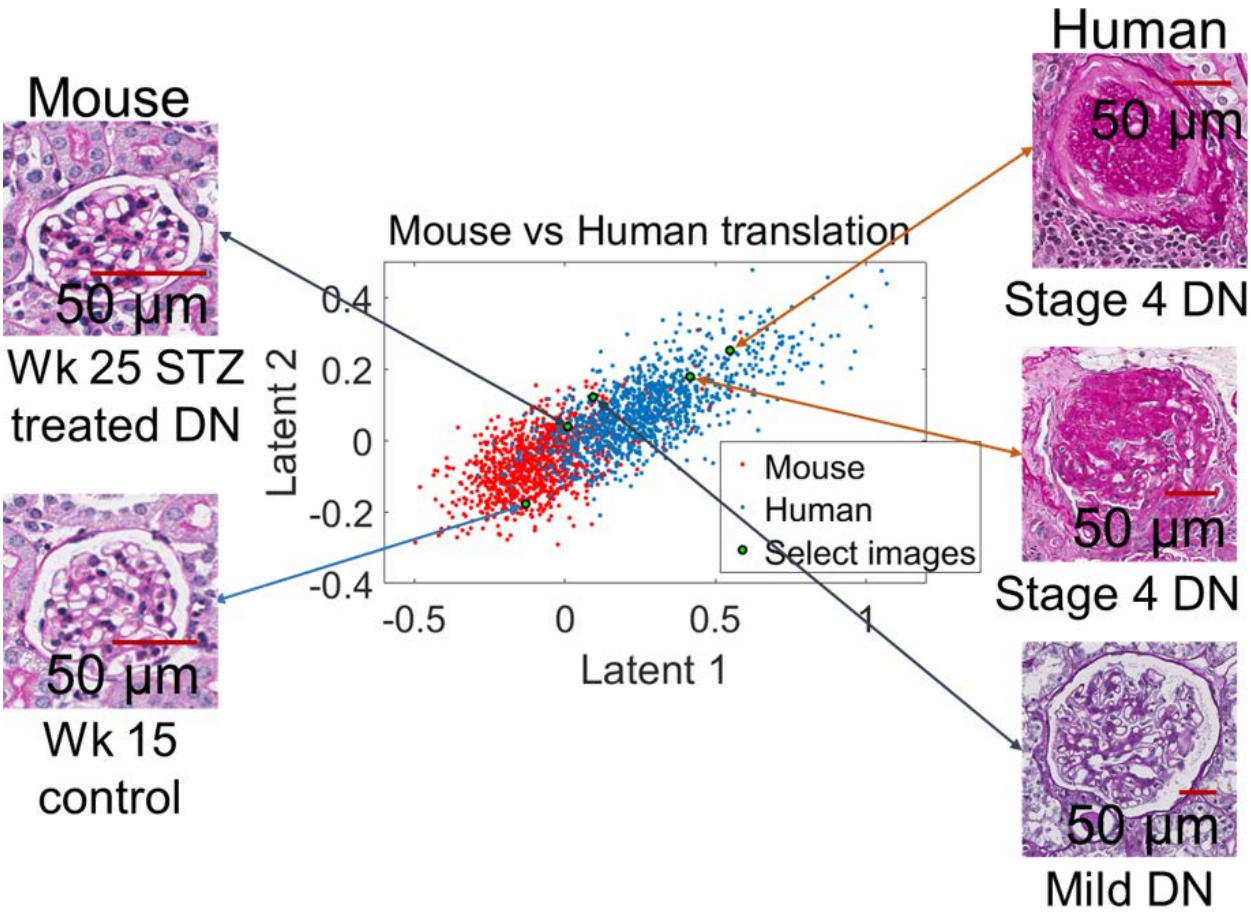


**Figure 3. Preliminary VAE Glomerular mapping results visualized, plotting 2K labeled glomeruli.** Trained using the architecture described in the network implementation section. A clear overlap between mouse and human data is noted with trends in DN progression observed. Human mild DN[7], and STZ treated mouse glomeruli[6] with late stage DN occupy the region of overlap, with mouse data absent in the upper right region where the most severe glomerular sclerosis occurs.

When selected glomeruli examples are encoded into the latent space and plotted, a trend in DN progression is visualized (Fig. 3). The authors note that the mouse model data used to generate Fig. 3 did not show that same level of glomerular sclerosis as the human biopsy data, with the most structural changes occurring in 25 week post STZ treated mice. This likely explains the distribution in the latent space, where 25 week post STZ treated mice overlap with mild DN human data-points.

## 4. DISCUSSION

We present (to our knowledge) the first attempt at automatic unsupervised feature determination of glomeruli, as well as the first cross species dataset in such an application. Despite the rudimentary state of our proposed network architecture, we have shown the viability of VAEs for modeling disease states in histopathological images.

Conceptually, autoencoders perform dimensionality reduction preventing one to one mapping of input data through various regulation techniques, which place constraints on latent dimensions. This constraint weighted against the reconstruction accuracy, ensures that input noise will not be encoded. In fact, denoising autoencoders purposefully inject noise into input nodes to ensure only the inherent data structure will be learned. VAE architectures use a similar technique, instead injecting noise into the central latent layer[9]. While noise modeling, is unlikely to affect cross species mapping using VAE's, data aberrations may affect latent mapping. For our data specimen preparation was conserved across species, wherever feasible, and currently, the preparation of our human and murine renal sections is identical, except for murine renal perfusion, which is not possible in human tissue.

In this implementation, mouse model data is used to augment limited human biopsy data, leveraging the complex automatic modeling of VAEs to statistically encode both datasets. DN progression, unlike other popular machine learning datasets, represents a continuous spectrum of structural themes to which stage labels are applied. As a result, the poor separability seen in Fig. 2B, showing the mapping of staged glomeruli into a 2-dimentional latent space, is unsurprising. This result is replicated in Fig. 3, where the latent representation is shown to encode information about the severity of DN structural changes.

The results in Fig. 2 show that data is clustered along a strong principal component axis in the latent space. We hypothesize that this was due to L2 regularization[10] used on the network parameters. While this has been shown in the literature to prevent overfitting, we will explore the relationship of L2 regularization and network architecture, particularly the number of latent features, removing it if we see fit. As a proof of concept, this method shows promising results with limited optimization, which will be improved in future iterations.

Perhaps the most interesting component of VAE architectures, is generative modeling. Once a network has been trained, the latent prior z can be sampled and decoded. This is unique to VAE architectures, and can be used to visualize the latent relationships learned by the network. We expect the ability to easily translate between spatial and latent data representations, where both spaces are defined by probability distributions, will enable unprecedented domain expert validation for further network training and testing. Currently we do not present results on generative decoding of latent variables due to the low dimensionality of our latent spaces. During testing we found that reconstructions from our 2-dimentional latent spaces were visually poor, however in future work we plan to extensively evaluate generative latent sampling using higher latent dimensions. Using this, we expect the ability to validate our network performance, using the domain expertise of our co-authors with clinical pathology expertise to visually examine relationships between decoded latent samples in future work.

## 5. FUTURE WORK

In future work we intend to replace the fully connected layers with sets of convolutional and pooling layers. Convolutional layers ensure spatial invariance to image features, while greatly reducing the number of trainable parameters in the network, enabling more complex architectures with higher dimensional inputs. We plan to explore the generative ability of such a network for semi-supervised data augmentation in other supervised models, as well as manifold learning for exploration of the latent feature space. Specifically we would like to utilize this network to motivate a transfer of knowledge across species, to better correlate features from animal models with human data. The overlap in mouse and human data predicted by our model (Figs. 2, 3) suggests that this data is indeed continuous, and therefore can be correlated statistically.

In future work, if aberration from data preparation is determined to adversely affect network performance, we will encode non-perfused murine data, comparing its latent encoding to a holdout perfused set. Given the network is affected by aberration we expect the non-perfused data set to overlap more with human data of similar preparation. This shift can be quantified using metrics such as average pointwise minimum Euclidean distance between the sets, for different latent dimensionalities. Due to latent enforcement over an i.i.d. prior, we expect in higher dimensional latent spaces, this aberration will only present itself in a few latent dimensions, with little effect on the latent representation overall. Latent dimensions affected by the aberration will be corrected, informed by the distance from the non-perfused murine data, shifting the perfused mouse data relative to the non-perfused mouse data.

## ACKNOWLEDGEMENT

This project was supported partially by the faculty start-up fund from the Pathology & Anatomical Sciences Department, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo (UB), partially by the UB IMPACT award, and partially by the DiaComp Pilot and Feasibility Program grant #32307-5. We thank NVIDIA Corporation for the donation of the Titan X Pascal GPU used for this research (NVIDIA, Santa Clara, CA).

## REFERENCES


[1] Bertram, J. F., Douglas-Denton, R. N., Diouf, B. *et al.*, "Human nephron number: implications for health and disease," Pediatr Nephrol, 26(9), 1529-33 (2011).

[2] Doersch, C., [Tutorial on Variational Autoencoders], (2016).

[3] Burda, Y., Grosse, R. and Salakhutdinov, R., [Importance Weighted Autoencoders], (2015).

[4] Hastie, T., Tibshirani, R., Friedman, J.H., [The Elements of Statistical Learning: Data Mining, Inference, and Prediction], Springer, 768 (2009).

[5] Numerical recipes in C the art of scientific computing Available from: http://www.nrbook.com/a/bookcpdf.php

[6] Tesch, G. H. and Allen, T. J., "Rodent models of streptozotocin-induced diabetic nephropathy," Nephrology (Carlton), 12(3), 261-6 (2007).

[7] Tervaert, T. W., Mooyaart, A. L., Amann, K. *et al.*, "Pathologic classification of diabetic nephropathy," J Am Soc Nephrol, 21(4), 556-63 (2010).

[8] Ginley, B., Tomaszewski, J. E., Yacoub, R. *et al.*, "Unsupervised labeling of glomerular boundaries using Gabor filters and statistical testing in renal histology," Journal of Medical Imaging, 4(2), 021102: 1-12 (2017).

[9] Jiwoong Im, D., Ahn, S., Memisevic, R. *et al.*, "Denoising Criterion for Variational Auto-Encoding Framework," ArXiv e-prints, 1511, arXiv:1511.06406 (2015).

[10] Ng, A. Y., [Feature selection, L1 vs. L2 regularization, and rotational invariance] ACM, Banff, Alberta, Canada(2004).