Evaluation of quantitative magnetic resonance imaging to predict intracranial hypertension in neonates by machine learning

Ishaani Khatri

Thesis Advisor: Dr. Ritambhara Singh

Second Reader: Dr. Harrison Bai

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science with Honors in the Computational Biology concentration of Brown University.

Evaluation of quantitative magnetic resonance imaging to predict intracranial hypertension in neonates by machine learning

Ishaani Khatri

April 2021

1 Abstract

The early and accurate diagnosis of neonatal intracranial hypertension (NICH) is clinical challenge, due to the lack of clinical symptoms in infants. Accurate diagnosis methods are invasive and involve some risk. Previous studies have shown that NICH may lead to specific clinical, morphological changes in the brain that can be detected through MRI scans.

Thus, there is need to develop data-driven, non-invasive methods to diagnosis NICH based upon MRI scans. Studies that apply machine learning to clinical imaging data are often hampered by the large number voxels and low number of subjects, known as the small-n-large-p problem. Feature reduction is vital step before training a machine learning model to mitigate the small-n-large-p problem.

In this study, we develop and evaluate the performance a deep learning-based autoencoder to feature reduce our data before applying machine learning meodels. We compare its performance to other popular feature reduction methods such as principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE).

2 Introduction

Elevated intracranial pressure (ICP), or intracranial hypertension, is a serious threat to the health of neonates, infants less than four weeks old. It is often caused by hypoxic-ischemic encephalopathy, bacterial meningitis or neonatal hypoglycemia [1]. The annual pediatric incidence of ICH is approximately 0.63 per 100,000 children [2]. Neonatal intracranial hypertension (NICH) can lead to negatively affect the neurological development of a child, impacting their long-term quality of life. More severe cases of NICH may compromise cerebral perfusion pressure and leading to herniation syndromes and death[3]. Thus, the

early detection of NICH is one of the keys to reducing neurological complications in neonates.

However, the diagnosis of NICH is quite difficult due to lack of clinical symptoms in infants[4-5]. To accurately measure intracranial pressure and correctly diagnose neonatal intracranial hypertension, a lumbar puncture procedure must be carried out. However, this procedure is invasive, and has a risk of hemorrhage and infection in infants. Thus, there is an urgent need to develop a non-invasive technology to detect neonatal intracranial hypertension.

Thus, developing a data-driven approach to predict NICH from MRI scans of neonates can be a promising direction towards non-invasively diagnosing infants in a timely manner. The application of machine learning techniques to clinical imaging data is increasingly being used in making relevant predictions on the individual patient level[6]. However, these clinical imaging studies are commonly hampered by the large number of features and a low number of observations, known as the small-n-large-p problem. To mitigate this problem, feature reduction is essential before applying of machine learning classifiers to clinical imaging data.

We propose using an autoencoder to feature reduce neuroimaging data from a 3D MRI scan. Unlike previous methods, such as PCA and t-SNE, an autoencoder is neural network-based feature reduction method. Therefore, it is able to to create a lower-dimensional representation of our data in a non-linear fashion. Thus, it may be able to capture the complexities of our 3D MRI imaging data more accurately.

3 Related Work

The recent development in MRI scans provides a way to obtain a precise quantitative assessment in disease prediction. However, their usability for accurate diagnosis of neonatal intracranial hypertension is yet to be seen.

Conventional imaging studies have been been previously conducted to identify intracranial hypertension in the adult population. These studies have shown that increased intracranial hypertension may lead to specific clinical, morphological changes in the brain that may be detected using non-invasive measurements like MRI scans. For instance, increased ICP may lead to changes of the ventricular system, optic nerve sheath diameter and pituitary gland[7-9]. However, the changes in these areas are usually secondary and are susceptible to the subjective experience and interobserver variability. Additionally, mild ICH doesn't cause morphological changes, so the application of these studies is limited.

More advanced studies have developed an algorithm to estimate ICP and diagnose ICH using cerebral blood flow[10]. Other groups have proposed predicting ICH by considering brain volume, cerebral spinal fluid, and cerebral blood flow. However, as these studies have predetermined features for ICH prediction, they may be missing out on other signs of intracranial hypertension. A machine learning approach is data-driven, and does not rely on domain knowledge. Thus, this methodology provide us with a non-invasive technology

to evaluate the MRI scan in its entirety in order to diagnose NICH early and accurately.

Studies that apply machine learning methods to clinical imaging data, including ours, are commonly hampered by the large number features (number of voxels in 3D MRI scan) and low number of subjects, known as the small-n-large-p problem or the curse of dimensionality. Feature reduction is a vital step before training a machine learning model. It mitigates the small-n-large-p problem, thereby avoiding overfitting and improving model prediction accuracy and generalization ability[6].

As discussed in a review article by Mwangi and his colleagues, in the field of neuroimaging, the most common methods of unsupervised feature reduction are principal component analysis and independent component analysis. Principal component analysis (PCA) can only learn linear transformation of the features. Therefore, it may not always capture the non-linear complexities that underly the features in clinical imaging data.

As the field of deep learning has expanded over the past few years, deep learning techniques are now able to be applied to feature reduction problems. The autoencoder is a neural network based technique that compresses high-dimensionality data into a low-diemsionality space, and then reconstructs the data back into the high-dimensionality space. Thus, the compressed feature space, called the bottleneck or latent space, contains a feature-reduced version of the input data. By extracting the data in the bottleneck space, we are essentially extracting a low-dimensionality representation of our data that was created in a non-linear fashion.

4 Methods

4.1 Imaging Data

Our data was obtained by collaborators at the Xiangya Hospital of Central South University, a Class-A Grade-3 general hospital located in Changsha, Hunan, China. The data was collected with the approval of the institutional ethics committee of the hospital. The Rhode Island Hospital institutional review board determined this project did not constitute as human subject research, as we received external, deidentified data.

Magnetic resonance imaging scans of the head were carried out on 85 neonatal patients from January 2017 to December 2019. MRI scans were conducted with the 3.0 T Siemens Prisma Human MRI scanner. Among those patients, 58 patients were diagnosed with neonatal intracranial hypertension (NICH), and 38 diagnosed to not have NICH. Patients were diagnosed by clinicians who performed lumbar punctures and measured their of the intracranial pressure.

Due to individual differences in human brains, the spatial coordinates of images in the scanning process are different. Therefore, individual differences must be eliminated in the first place to unify the coordinates. That is, all the brains of the subjects need to be corrected, or registered, on the standard

template, so that subsequent statistical analysis can be conducted. Domain experts manually completed this process, providing us with registered images for each patient. The ground truth annotation labels were as following:

Label 0 Patients without NICH

Label 1 Patients with NICH

Figure 1: Flowchart demonstrating data collection and split.

4.2 Data Preprocessing

Before inputting our images into our feature reduction steps, every MRI image was converted from Neuroimaging Informations Technology Initiative (NIFTI) format to a NumPy array (https://numpy.org). The MRI images were three-dimensional head and neck scans of size size 117 x 159 x 126. Each 3D voxel was of size 0.78 mm x 0.78 mm x 0.78 mm.

These three dimensional images were flattened into one-dimensional arrays. Thus, they were able to feature-reduced by the t-distributed Stochastic Neighbor Embedding (t-SNE) and Principal Component Analysis (PCA) methods.

To utilize our autoencoder, which expects two-dimensional inputs, we separated each three-dimensional image into slices by its z-axis. Thus, all 85 three-dimensional images of size 117 x 159 x 126 were separated into 10710 two-dimensional images of size 117 x 159.

4.3 Feature Reduction Methods

4.3.1 Principal Component Analysis

As the first baseline method, we utilized principal component analysis (PCA) for feature reduction. PCA reduces dimensionality of data while retaining most of the variation in the dataset by identifying directions along which the variation in the data is maximal [12]. These directions, called principal components, now represent the variation of the data in a lower dimensionality space. By using a few components, each patient sample can be represented by relatively few numbers instead of by values for thousands of variables. Of note, PCA can only learn linear transformation of the features.

We applied PCA by utilizing a principal component analysis implementation from scikit-learn[13]. Our original images contained 2343978 features. These images were reduced to their top 10, 20, 30, 40, 50, 60, 70, and 80 principal components. All of these eight PCA-reduced datasets were now ready to be input into our machine learning models. We initially hoped to also reduce our datasets into their top 90 and 100 principal components as well. However, the implementation of PCA we used allowed for us to reduce our data to a maximum of 85 components, since the dataset contained 85 samples.

4.3.2 t- Distributed Stochastic Neighbor Embedding

As an additional baseline method, we utilized t-Distributed Stochastic Neighbor Embedding (t-SNE) for feature reduction. Unlike PCA, t-SNE is is a non-linear dimensionality reduction algorithm [14]. It maps multi-dimensional data to two or more dimensions suitable for human observation.

We applied t-SNE by utilizing a principal component analysis implementation from scikit-learn[13] .As a part of this process, we had to tune the perplexity hyperparameter of t-SNE. After visualizing our data in two dimensions with perplexity values ranging from 5 to 50, incremented by 5, we found that a perplexity value of 45 worked best for our data.

Though t-SNE is commonly used to reduce data to two or three dimensions for visualization purposes, we utilized it to reduce our 2343978 feature images to 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 components. These ten t-SNE reduced datasets were now ready to be input into our machine learning models.

4.3.3 Autoencoder

In our study, we built an autoencoder in order to reduce the dimensionality of our data. An autoencoder has two parts: the encoder and the decoder. The encoder reduces the high-dimensionality image into a low-dimensional space, called the bottleneck or latent space. From the latent space representation, the decoder then reconstructs the image [15].

Of note, our autoencoder took in 10710 two-dimensional images of size 117 x 159. The 10710 two-dimensional images were extracted by slicing the 3D MRI images from all 85 patients into 126 two-dimensional images. Our encoder included the following: a convolutional layer, a max pooling layer, a second convolutional layer, a second max pooling layer, a third convolutional layer, and the final dense layer. Our decoder, essentially a reversal of the encoder, included the following: a dense layer, a convolutional layer, an upsampling layer, a second convolutional layer, a second upsampling layer, and a third convolutional layer.a The autoencoder took in 10710 two-dimensional images of size 117 x 159.

After the autoencoder was trained, we were able to extract the latent space representation. We performed this process with a specified latent space of size 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100. Thus, our 2D images of size 18603 were reduce to these sizes. Then, all the latent size representations from the

same sample were concatenated into a single row, creating feature array with the 85 samples and 126*latent-size features.

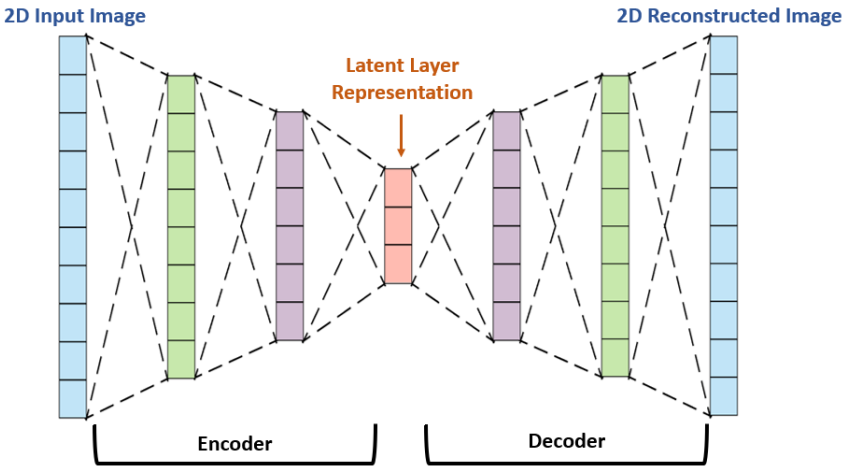

Figure 2: Autoencoder architecture.

4.4 Classification Methods

With our datasets now considerably reduced, we were ready to input our datasets into the machine learning algorithms. Given the small size of our dataset, we decided to use two relatively simple binary classification algorithms: logistic regression (LR) with regularization and support vector machines (SVM).

Both of these were implemented via implementations from scikit-learn [13]. Of note, we utilized a support vector machine model with a linear kernel and another support vector machine with a rbf kernel.

4.5 Evaluation

Given our small dataset size, we evaluated our data with a 3-fold training and testing scheme. Our data was divided into 3-folds, with two of the folds being utilized for the training and the third held out as our testing set. The third that was selected as the testing set was then changed, and our model was retrained and tested. This was repeated three times. Thus, the resulting output was three of every metric, of which we found the average and standard deviation.

We evaluated our models with the following metrics: average precision score and area under the receiver operating characteristic curve (ROC AUC). These metrics take in account the slight class imbalance of our dataset.

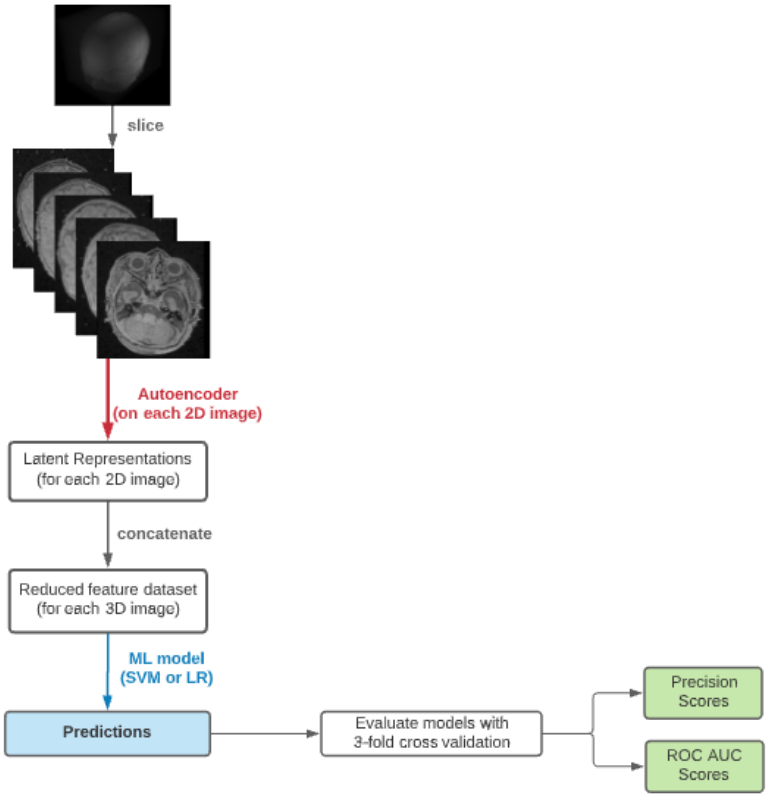

Figure 3: Methods overview.

5 Results

An important result of this study was the development of our feature reduction and machine learning pipeline to predict NICH.

Training our autoencoder was an important and computationally-expensive step to feature reduce our input data. In this step, the layers in the decoder and encoder of the autoencoder were tuned in an attempt to minimize the loss between the original image and reconstructed image. As Figure 4 shows, the size of the latent layer representation did not drastically affect the loss between the original image and the reconstructed image. After 10 epochs, the loss across the different size latent layer representations remained relatively close to one another.

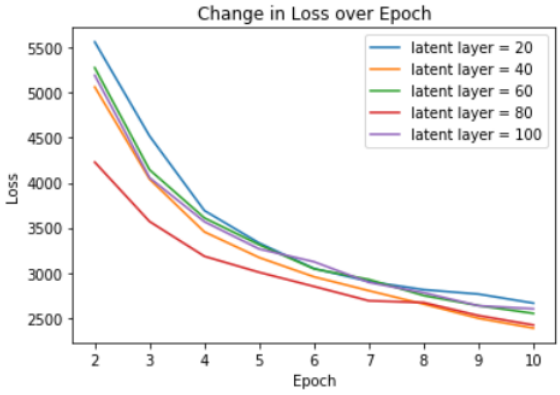

Figure 4: Change in loss for autoencoders of various latent layer representation sizes.

After feature reduction, we applied three machine learning models to our newly reduced data: 1) a logistic regression classifier 2) a support vector machine classifier with a linear kernel and 3) a support vector machine classifier with an rbf kernel. Of the three methods, the SVM with an rbf kernel was most successful in classifying our data, as shown by its high precision scores and ROC AUC scores [Figure 5, Figure 6]. This is likely because this method was able to capture the non-linear decision boundary between the data points of the two classes. For the sake of visualization, we will compare the performance of these three methods to predict NICH based on the same autoencoder-reduced dataset.

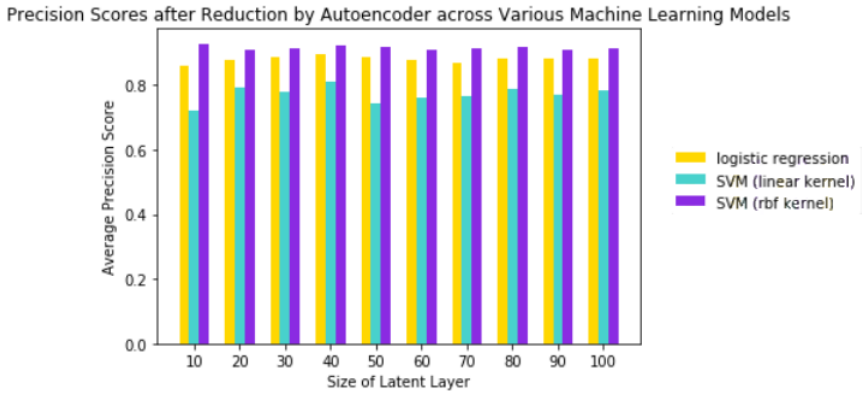

Figure 5: Precision scores after feature reduction by autoencoder at various latent layer sizes for three machine learning models: logistic regression, SVM with a linear kernel, and SVM with an rbf kernel.

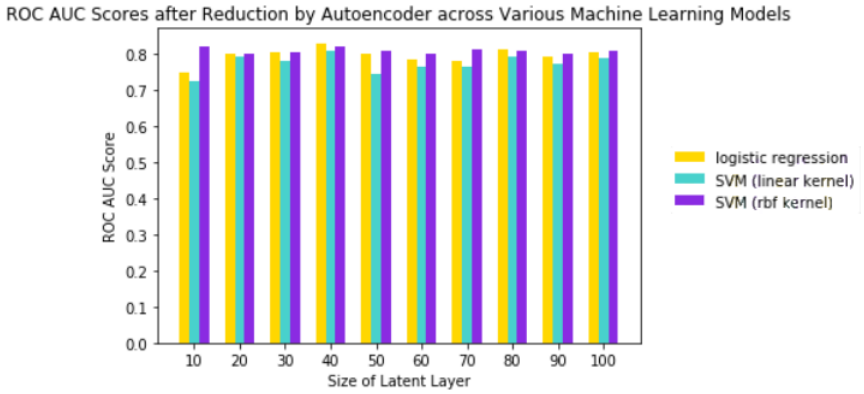

Figure 6: ROC AUC scores after feature reduction by autoencoder at various latent layer sizes for three machine learning models: logistic regression, SVM with a linear kernel, and SVM with an rbf kernel.

With the SVM with an rbf kernel selected as the best machine learning model, we now compared the performance of this machine learning model to classify our data based on the three datasets: one reduced by an autoencoder, one reduced by PCA, and one reduced by t-SNE. Our evaluation demonstrates that, after being input into the same machine learning model, the dataset reduced by an autoencoders consistently performed better in this prediction task than the datasets reduced by PCA and t-SNE, as shown by the higher ROC AUC and precision scores [Figure 7, Figure 8]. Of note, for the size of the latent layer parameter of the autoencoder, this value is per 2D slice for the MRI. For PCA and t-SNE, it is per 3D MRI image. Thus, the autoencoder reduced feature matrix was larger in size than than PCA and t-SNE reduced feature matrices.

As different feature reductions methods performed optimally at different latent sizes, the most fair comparison of the three methods is to compare their best performing metrics, which may occur at varying latent representation sizes. Comparing these best performing metrics, we demonstrate that the data reduced by the autoencoder continued to achieve the higest precision score of 0.928 and the highest ROC AUC score of 0.820[Figure 9, Figure 10].

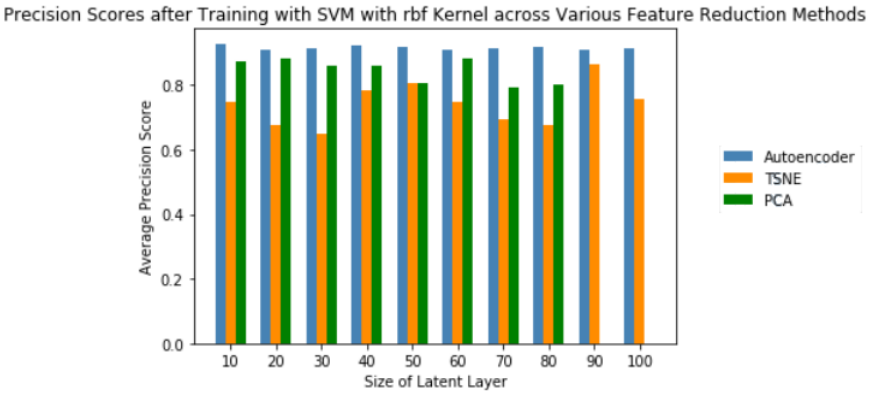

Figure 7: Precision scores after feature reduction by autoencoder, t-SNE, and PCA at various latent layer sizes and utilizing the same machine learning model, a SVM with an rbf kernel.

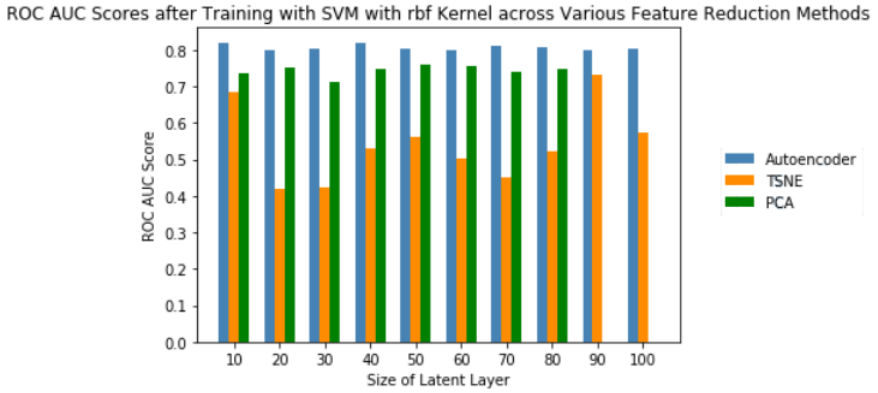

Figure 8: ROC AUC scores after feature reduction by autoencoder, t-SNE, and PCA at various latent layer sizes and utilizing the same machine learning model, a SVM with an rbf kernel.

To conclude, the best performance, with respect to both precision and ROC AUC score, was the achieved by applying an SVM with an rbf kernel to data reduced by the autoencoder with a latent layer representation of size 10.

6 Discussion

Our results have shown that autoencoders can effectively feature reduce large clinical images, preparing them for training by machine learning models such as SVMs and logistic regressions. The data reduced by autoencoders performed better with respect to precision scores and AUC scores compared to data reduced

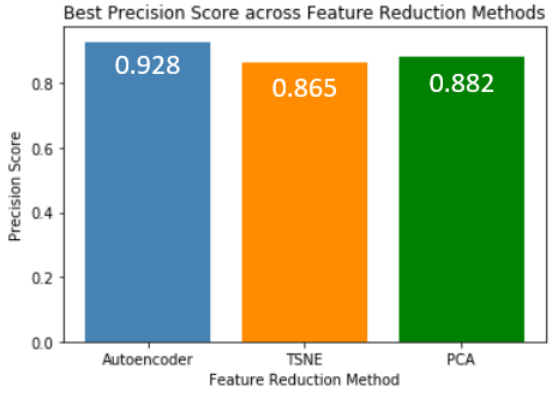

Figure 9: AUC scores after feature reduction by autoencoder, t-SNE, and PCA at various latent layer sizes and utilizing the same machine learning model, a SVM with an rbf kernel.

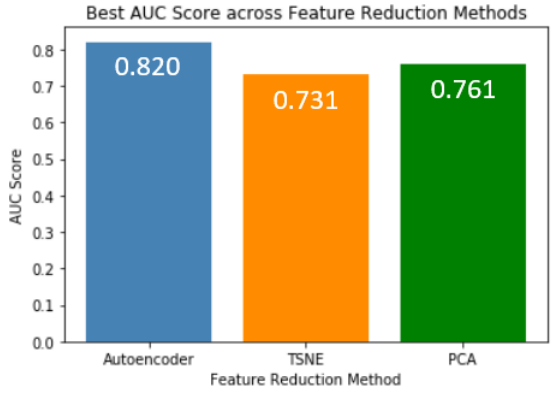

Figure 10: ROC AUC scores after feature reduction by autoencoder, t-SNE, and PCA at various latent layer sizes and utilizing the same machine learning model, a SVM with an rbf kernel.

by PCA and t-SNE. This is particularly noteworthy, as these two methods, particulary PCA, are commonly used in the field of neuroimaging to reduce features.

These results are very promising, showing that feature reduction by autoencoders provides reduced dimensionality datasets that are informative for machine learning models. Despite our small sample size, our models were able to output predictions with relatively high precision scores and AUC scores. We were impressed by the success of our autoencoder feature reduction and machine learning models.

The greatest limitation of our study was our small sample size. Due to the small sample size, we were unable to have a validation dataset, and had to utilize a scheme that divide our data into only training and test datasets. It is also possible that our small sample size may have resulted in some overfitting.

In the future, we hope to apply our autoencoder feature reduction methodology to other clinical imaging datasets, and learn if it can be successfully applied to answer other clinical imaging-based classification problems. Additionally, we hope to tune the hyperparameters and architecture of our autoencoder. We would also like to attempt an approach that utilizes 3D convolutional layers.

To evaluate our autoencoder feature reduction method further, we would also like to apply it to larger datasets. With larger datasets, we could apply more machine learning algorithms. We could also take a fully deep learning based approach, and apply neural network to clinical imaging data.

7 Acknowledgments

I would like to thank Dr. Singh for providing me with the opportunity to work in her lab and for her constant support throughout this project. Dr. Singh has been an incredibly considerate and supportive mentor, and her support and mentorship have been invaluable in my growth as both a student and computational biologist.

I would also like to thank Dr. Bai for his support as the clinical mentor of this project. He and his team provided us with an exciting dataset to work with, as well as clinically relevant information for this project. I appreciate the time and patience of Rong Hu and Jing Wu, who were so helpful in explaining the clinical imaging dataset. I am so grateful for the opportunity to work on this interesting, clinically-focused computational project.

Lastly, I would like to thank the Singh Lab as a whole. Everyone in the lab has been so warm and generous with their time, and so willing to help one another out. I especially appreciate the patience and help of Adam Berkley and Sally Zhi. I am very lucky to have been a member of such a great group!

8 References

1. Sands R, Shanmugavadivel D, Stephenson T, Wood D. Medical problems presenting to paediatric emergency departments: 10 years on. Emerg Med J. 2012;29(5):379-382.

2. Gillson N, Jones C, Reem RE, Rogers DL, Zumberge N, Aylward SC. Incidence and Demographics of Pediatric Intracranial Hypertension. Pediatr Neurol. 2017;73:42-47.

3. Welch, Timothy P et al. Fentanyl and Midazolam Are Ineffective in Reducing Episodic Intracranial Hypertension in Severe Pediatric Traumatic Brain Injury. Critical care medicine vol. 44,4 (2016): 809-18.

4. Le Roux P. Intracranial Pressure Monitoring and Management. In: Laskowitz D, Grant G, editors. Translational Research in Traumatic Brain Injury. Boca Raton (FL): CRC Press/Taylor and Francis Group; 2016. Chapter 15. Available from: https://www.ncbi.nlm.nih.gov/books/NBK326713/

5. Raboel, P H et al. Intracranial Pressure Monitoring: Invasive versus Non-Invasive Methods-A Review. Critical care research and practice vol. 2012 (2012): 950393. doi:10.1155/2012/950393

6. Mwangi, Benson et al. A review of feature reduction techniques in neuroimaging. Neuroinformatics vol. 12,2 (2014): 229-44. doi:10.1007/s12021-013-9204-3

7. Ganslandt O, Mourtzoukos S, Stadlbauer A, Sommer B, Rammensee R. Evaluation of a novel noninvasive ICP monitoring device in patients undergoing invasive ICP monitoring: preliminary results. J Neurosurg. 2018;128(6):1653-1660.

8. Huang-Link Y, Eleftheriou A, Johansson JM, et al. Optical coherence tomography (OCT) Represents a sensitive, reliable tool for routine monitoring patients with idiopathic intracranial hypertension (IIH). J Neurol Sci. 2017;381S:149.

9.Hu R, Holbrook J, Newman NJ, et al. Cerebrospinal Fluid Pressure Reduction Results in Dynamic Changes in Optic Nerve Angle on Magnetic Resonance Imaging. J Neuroophthalmol. 2018.

10. Kashif FM, Verghese GC, Novak V, Czosnyka M, Heldt T. Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure. Sci Transl Med. 2012;4(129):129r-144r.

11. Velliangiri S, Alagumuthukrishnan S, Thankumar SI. A Review of Dimensionality Reduction Techniques for Efficient Computation, Procedia Computer Science, Volume 165, 2019. https://doi.org/10.1016/j.procs.2020.01.079.

12. Ringnér, Markus. What is principal component analysis?. Nature biotechnology vol. 26,3 (2008): 303-4. doi:10.1038/nbt0308-303

13. Pedregrosa et al. Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830, 2011.

14. van der Maaten L, Hinton G. Visualizing Data using t-SNE. Journal of Machine Learning Research 9 (2008) 2579-2605.

15. Vincent, Pascal, et al. "Extracting and Composing Robust Features with Denoising Autoencoders." Proceedings of the 25th International Conference on Machine Learning - ICML '08, 2008, doi:10.1145/1390156.1390294.