CS230

Prediction of postoperative opioid continuation in patients undergoing lumbar fusion

Chloe O'Connell
Stanford University School of Medicine
chloeo@stanford.edu

Felipe Kettlun
Stanford Computer Science
fkettlun@stanford.edu

Thomas A. Petersen
NVIDIA/SCPD
tap22sf@stanford.edu

Abstract

Given the addictive properties of opioid medications, they are not recommended after the immediate recovery period following surgery. However, approximately 20% of patients undergoing lumbar fusion meet criteria for "chronic opioid use" more than 3 months after surgery [1]. Despite growing awareness about the opioid epidemic sweeping the US, we do not have currently any way to predict which patients will continue taking opioids for prolonged periods of time following surgery. In this project, we attempted to predict which patients were at risk for continued opioid usage after 3 months following lumbar fusion using administrative healthcare data. After conversion of diagnosis codes into their corresponding 300-dimensional encodings, an alternative training dataset was also created using principal components analysis (PCA) to reduce the dimensionality of these encodings into their most informative two dimensions. Ultimately, our model was able to achieve similar predictive accuracy on these two datasets, with AUCs of 0.725 and 0.728 for the raw encodings and dimensionality-reduced datasets, respective. This network performed as well or better than logistic regression (AUC = 0.721) and random forest (AUC = 0.724) classifiers trained on the same dataset.

1 Introduction

The opioid epidemic has garnered increasing attention in the past few years, in both the news media and the medical literature (Haffajee et al., 2017; Nelson et al., 2015; Compton et al., 2016). Between 2000 and 2009, the number of outpatient prescriptions for opioids doubled (Governale et al., 2010), accompanied by a parallel rise in overdoses. Opioid related overdose deaths increased 200% between 2000 and 2014 (Rudd et al., 2016), reaching an all-time high of 33,091 in 2015, marking a four-fold increase compared to 1999 (O'Donnell et al., 2017). By 2015, opioids accounted for 63% of all fatal drug overdoses in the United States, with approximately half of these involving prescription opioids (Rudd et al., 2016). Although opioids play a useful role when it comes to control of postoperative or cancer-related pain, even prescription opioid use under the supervision of a physician has been associated with increased risk for opioid use disorders. A patient who receives a prescription for opioid medication is thought to have anywhere up to a 122-fold increase in the risk of the opioid abuse or dependence compared to one who has never received an opioid prescription (Edlund et al., 2014). Even though opioids are not recommended for chronic pain, the medical use of opioids is unfortunately common in patients with low back pain. As of 2015, opioids were the most common class of drugs prescribed for this condition (Deyo et al., 2015). For patients with particularly severe structural disease, lumbar fusion can occasionally help alleviate the need for these addictive medications. However, surgery is not a perfect fix, and approximately 20% of patients undergoing lumbar fusion meet criteria for "chronic opioid use" more than 3 months after surgery (O'Connell et al., 2017). Despite growing awareness about the opioid epidemic sweeping the United States, we do not have any way to know which patients will continue taking opioids for a prolonged period following surgery.

In this project, we attempted to predict which patients were at risk for continued opioid usage more than 3 months following lumbar fusion using only administrative healthcare data which would be available before surgery and could if accurate help guide physician decisions.

2 Related work

Successes with deep learning in the medical domain have been largely restricted to imaging (Gulshan et al., 2016; Zhang et al., 2015; Rajpurkar et al., 2017). A variety of characteristics of medical data have hindered the success of deep learning in the medical domain, including the lack of standardization of electronic medical records between hospitals, HIPAA privacy laws preventing open-source access to data, and the diverse way information is stored in clinical notes. With the growing popularity of sequence models and advancements in natural language processing, deep learning has begun to experience greater success in healthcare fields outside of imaging. Nguyen et al. (2017) used a convolutional neural network to capture both time-dependent and independent aspects of medical records and capture local clinical motifs to stratify patient risk. A combination of both supervised and unsupervised layers in the form of a Restricted Boltzmann Machine has also been applied to ICU data, out-performing both support vector machine and gradient boosted machine models when it came to predicting 28-day mortality (Du et al., 2016).

However, in the case of more structured data (i.e. administrative data collected by healthcare insurance agencies with predefined features), neural networks have generally not been able to outperform support vector machine (SVM) or random forest approaches (Eskidere et al., 2012; Dinov et al., 2016). This is likely because much of the richness of structured healthcare data lies in the ICD-9 (now ICD-10 as of 2017) diagnosis codes assigned to each patient, which present multiple problems to machine learning algorithms. First, there are varying numbers of ICD-9 codes associated with each patient and inconsistent code usage between physicians. Second, there are 13,000 ICD-9 codes (and a full 68,000 in the updated version of this system, the ICD-10). Although there is some biological relatedness between many ICD-9 codes, there is no innate structure to the ways these diagnosis codes are organized. For example, there are 12 different ICD-9 codes that correspond to some sort of stroke, and yet no current way to link related biological concepts to each other except a manual search and literature review. With the lack of structure in the way this data is encoded, it is unsurprising that much of the deep learning research in the non-imaging healthcare domain data has gone into discovering useful representations of these codes. In 2016, Choi et al. (2017) used over 4 million electronic healthcare records to identify concept embeddings of ICD-9 codes from administrative claims data. These encodings mapped the ICD-9 codes into 300-dimensional space in which related codes were near each other in terms of Euclidean distance. These encodings better represent medical relatedness compared to ICD-9 codes in a way that is more easily interpretable by machine learning algorithms.

One exception to the relatively disappointing performance of neural networks compared to RF and SVM was work done by Jamei et al. (2017), who used a shallow network (two dense fully connected layers) and was able to slightly outperform random forest (validation AUC of 0.78 vs. 0.77) when predicting unplanned hospital readmissions within 30 days. However, they were assisted by predefined categories that were automatically computed by the Sutter Health Electronic Medical Record System and therefore did not need to perform any other post-processing on patients' associated ICD-9 codes. Given their success with a relatively shallow network, our team decided to determine whether a similarly shallow neural network could outperform other methods with the addition of diagnosis encodings identified by Choi et al. (2017).

3 Dataset and Features

Inclusion criteria: The final dataset consisted of 140,848 patients enrolled in employer-provided health insurance plans with data collected by the Truven MarketScan database. These patients underwent lumbar fusion between 2007 and 2013, and were enrolled in their health insurance plan for at least 6 months prior to surgery and at least 1 year following surgery to ensure adequate follow-up.

Predictors: Our dataset included information on patient demographic factors (age, sex, geographic location), as well as a variety of relevant comorbidities that had been pre-computed based on ICD-9 and CPT codes present in other linked databases in the months preceding surgery (Table 1). The pre-defined categorical geographic variables in this dataset (i.e. state) were augmented by a computed longitude and latitude value for each patient, calculated based on the center of the patient's metropolitan statistical area. Geographic location as a combination of continuous numeric variables was desired to have a more flexible way of capturing hotspots of opioid use in the United States, as these usage patterns do not necessarily follow state or zip code borders. Preoperative prescription drug information (including opioid usage and amount, as well as other psychotropic drugs) was extracted from a linked pharmacy database using a summation of each patient's prescriptions in the six months prior to surgery. Each prescription was converted to the milligram morphine equivalents (MMEs) using the conversion table provided by the CDC (CDC.gov).

Diagnosis code processing: In addition to the preprocessed comorbidities extracted from previous visits, our dataset also contained up to 15 ICD-9 codes per patient representing diagnoses associated with the hospital admission in which the patient underwent their surgery. To determine whether this information would help our predictive model, we generated two different datasets and compared predictive results between the models trained on these two datasets. First, we selected the "primary diagnosis" for each patient and substituted the 300-dimensional encodings of medical entities published by Choi et al. (2016). Given that only a subset of these learned encodings were present in our dataset, we also generated a second dataset which utilized PCA to reduce the feature space of the encodings of each patient's top two diagnoses from 300 dimensions each to 2 dimensions each (selected based on the two principal components that captured the most variability of all encodings in our dataset). A second neural network was trained on this "Principal Components" input data which substituted these four features (two per diagnosis code) in place of the patient's primary ICD-9 code.

Finally, numerical variables were scaled to a mean of zero and a standard deviation of one. Variables that were factors were converted into one-hot encodings.

Outcome: The primary outcome of interest was "opioid continuation", defined as the patient continuing to fill opioid prescriptions after the immediate preoperative period (months 3-12 following surgery).

Training and Validation: Our dataset was randomly shuffled and divided into training, development, and test sets using an approximate 90%/5%/5% split consisting of 126446, 7042, and 7360 patients respectively.

Feature (n = 17)	Feature type
Demographic information	
Age	Numeric
Year	Numeric
Total preoperative supply of opioids (accounts for dose)*	Numeric
Number of days' patient was prescribed opioids in 6 months prior to surgery	Numeric
Approximate geographic location of residence	
Longitude	Numeric
Latitude	Numeric
Geographic region of hospital	Categorical
Employee-provided health plan	Categorical (binary)
Sex	Categorical (binary)
Procedure characteristics	
Multiple-level laminectomy	Categorical (binary)
Multiple-level fusion	Categorical (binary)
Instrumentation	Categorical (binary)
Preoperative prescription information	
Preoperative antidepressant use	Categorical (binary)
Preoperative antipsychotic use	Categorical (binary)
Preoperative benzodiazepine use	Categorical (binary)
Preoperative total milligram morphine equivalent supply)	Numeric
Preoperative total days of opioid prescription	Numeric

Existing comorbidities (n = 22) (All binary categorical variables)	
Hypertension	Rheumatological disease
Congestive heart failure	Hemiplegia
Myocardial infarction	HIV/AIDS
Diabetes (untreated)	Peptic Ulcer Disease
Diabetes (treated)	Valve Disease
Obesity	Osteoporosis
Peripheral vascular disease	Cerebrovascular disease
Alcohol abuse	Dementia
Drug abuse	Chronic Obstructive Pulmonary Disease
Tobacco use	Mild liver disease
Treatment for tobacco addiction	Moderate/severe liver disease
Depression*	Metastatic cancer
Bipolar disorder*	Cancer (non-metastatic)
Schizophrenia*	Lymphoma
Anxiety*	Hypothyroidism
Pulmonary circulatory disease	
Neurodegenerative disease	

Table 1: Input feature set

4 Methods

The selected model utilized a fully-connected architecture with ReLu activation and binary cross-entropy loss function. This model architecture was selected based work by Jamei et al. (2017), who found that a shallow fully connected network was able to outperform a random forest classifier in predicting 30-day hospital readmissions. Because there is no "gold standard" for this classification that allows us to quantify Bayes' error, we chose to compare our neural network prediction to four different existing methods. First, we developed a simple classifier that predicted continued postoperative usage for every patient who had non-zero opioid usage in the 6 months prior to surgery. We next trained the logistic regression, random forest, and SVM classifiers using all features of the principal components dataset and evaluated their accuracy on the test set. Random forests and SVM were tuned using a variety of different hyperparameters and the best-performing method was selected. In the case of the SVM, computational limitations prevented us from training the classifier on the entire dataset, and therefore the training dataset had to be limited to 50,000 patients.

5 Experiments

Since the search space for a deep-learned network is extremely large we decided on a tiered strategy to find an optimal model. First, we focused on finding a suitable architecture, then we tuned hyperparameters of the model in order (learning rate, batch size, L2 regularization, and dropout). Our initial search used the full 300-dimensional encoding version of our dataset.

Architecture: We began by setting reasonable default hyperparameters for FC networks and variations of architecture. We focused our search on networks that had N nodes in the first hidden layer and N*2 in subsequent hidden layers. Supplemental Table 1 of the Appendix lists select variations of layer number (L) and nodes/layers (N) that were evaluated along with training and dev results. Figure 1 shows how the selected architecture was ultimately able to fit the training set over many epochs of training. Three hidden layers of 500/1000/1000 nodes per layer was sufficient to fit our dataset.

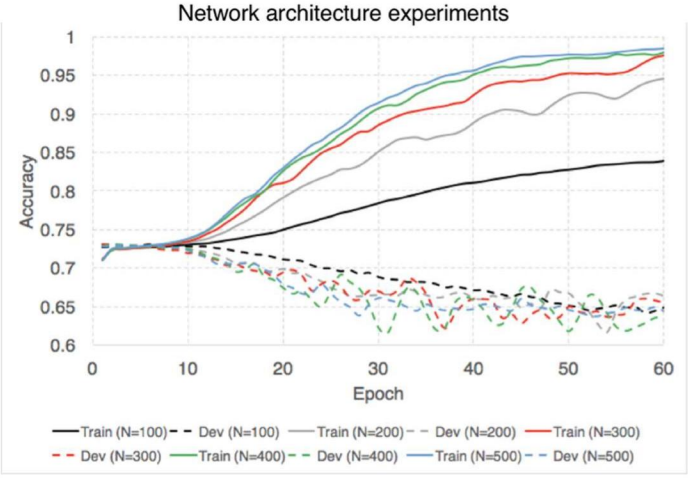

Figure 1: Layer size selection. N represents the number number of neurons in the first hidden layer, with each subsequent hidden layer having 2*N neurons.

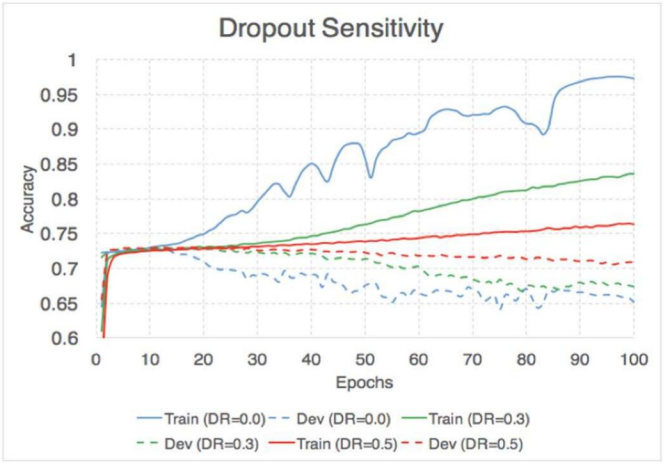

Figure 2: Dropout rate hyperparameter tuning

Hyperparameters: Given that dev set accuracy was very sensitive to changes in learning rate, this was the first hyperparameter we chose to tune. Figure 3 shows how our accuracy during training was dramatically impacted by learning rate. We chose the highest learning rate that was stable (i.e. the loss did not begin to diverge) which for our model and dataset was 1.0×10^{-3}. For batch size, our goal was to find the most computationally efficient batch size that delivered good accuracy. We found that a batch size of 8192 worked well on our GPU accelerated training. Per epoch training time was very fast, and the accuracy per epoch was nearly identical to lower batch sizes (Supplemental Table 2).

Finally, we focused on overfitting. We evaluated L2 regularization and varying levels of dropout and found that a dropout rate of 0.5 produced our best result. As expected during our sensitivity testing, we found that our training accuracy declined significantly as we increased dropout. Unfortunately, our best model was still only able to predict 72% on our dev set. Figure 2 shows the impact of dropout on our training and dev accuracy over epoch.

Our final model architecture is shown in Figure 4 which we ran with learning rate of 1.0×10^{-3}, batch size of 8192 and dropout of 0.5. For the Principal Components dataset, the same process was repeated with almost identical behavior. The optimal architectures were the same between datasets, with the only difference between hyperparameter selection being the choice of a dropout rate of 0.45 for the Principal Components dataset.

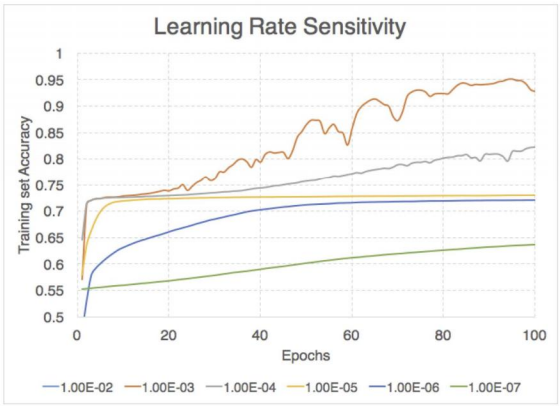

Figure 3: Learning rate hyperparameter tuning

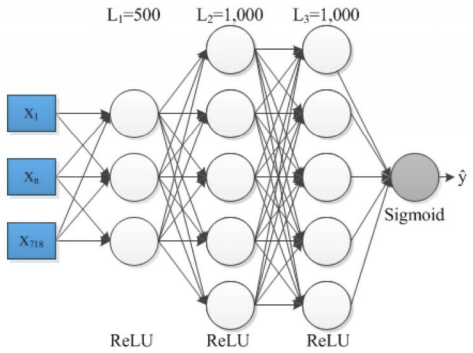

Figure 4: Final network architecture

6 Results/Discussion

Overall, similar sets of hyperparameters were optimal for networks trained on both the Principal Components and full Encodings datasets. The fully trained neural network using tuned hyperparameters was able to either equal or slightly outperform all of our comparison methods. On our test set, our best models achieved AUCs of 0.728 and 0.725 on the Principal Components and Encodings datasets, respectively (Table 2). The best-performing comparison method was the random forest classifier, with an AUC of 0.721. Although this AUC was more or less comparable to that of the neural networks, the NNs surpassed the random forest when it came to correctly identifying patients who would not continue their opioid use (recall of 0.833-0.841 for the neural networks versus 0.711 for the random forest).

	Test set metrics			
Model*	Accuracy	AUC	Precision (aka PPV)	Recall (aka NPV)
Simple binary (based on preop usage)	0.639	0.657	0.567	0.774
Logistic regression	0.721	0.778	0.740	0.711
Random Forest	0.724	0.775	0.737	0.717
SVM**	0.602	0.699	0.670	0.590
Neural network (PCs)	0.728	0.786	0.759	0.841
Neural network (Encodings)	0.725	0.787	0.751	0.833

PPV = positive predictive value, NPV = negative predictive value

*Comparison methods were conducted with the principal components dataset

** Due to computational constraints, the SVM was only trained on 50,000 patients rather than the full training set

7 Conclusions/Future Work

Overall, our neural network was able to slightly outperform existing methods predicting which patients would continue to use opioids more than 3 months after lumbar fusion.

Future work: Given time and computational constraints, we used a pre-processed form of the MarketScan database which had only one row per subject that contained information on comorbidities of interest in the year prior to surgery. However, the raw form of this database contains hundreds of rows per subject across multiple sub-databases (one for pharmacy information, another for outpatient visits, and another for inpatient hospital stays). To take advantage of the time series nature of this data, exploring a more complex recurrent network that can consider multiple patient visits as sequence data of varying lengths could be a potentially interesting approach. Although deep learning has not consistently outperformed other methods when it comes to structured health record data, semi-supervised approaches (including Restricted Boltzmann Machines, or RBMs) have shown some promise in making predictions based on healthcare record data from the intensive care unit. Further exploration of unsupervised learning approaches to data preprocessing, such as the unsupervised input of diagnosis codes into a RBM, would have been of great interest. Finally, training the network on two separate training datasets stratified by preoperative opioid usage could be an interesting next step, as there may be a clinical difference between patients who initiate opioids with surgery and are unable to stop versus patients for whom surgery is unable to alleviate a pre-existing opioid use habit.

8 Contributions

Thomas: Majority of the code writing and training. Ran hyperparameter tuning due to exceedingly helpful personal computational resources, Experiments section of final write-up.

Felipe: Wrote code for tuning architecture (loop through hyperparameters, train the model and save the results for each combination to a csv file, summary script generation (read csv files of each run and merge them to have only one table for comparing different runs), created a representative fake dataset that can be uploaded to GitHub.

Chloe: Initial idea generation and dataset procurement, dataset merging and cleaning (this part only was for another project), opioid use endpoint generation (had to be changed from previous opioid use endpoint that was in the data based on patients' prescription habits), data normalization and imputation of missing fields, latitude/longitude computation based on MSA or zip code, encodings substitution for ICD-9 codes and subsequent PCA, comparison methods, calculation of AUC/precision/recall for all methods, AUC figure generation, wrote up all sections of paper except experiments section.

9 References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. https://www.tensorflow.org/.

Choi Y, Yi-I Chiu CM, Sontag D. Learning Low-Dimensional Representations of Medical Concepts. http://people.csail.mit.edu/dsontag/papers/ChoiChiuSontag_AMIA_CRI16.pdf. Accessed May 18, 2018.

Compton WM, Jones CM, Baldwin GT. Nonmedical Prescription-Opioid Use and Heroin Use. *N Engl J Med*. 2016;374(13):1295-1296. doi:10.1056/NEJMc1601875.

Deyo RA, Von Korff M, Duhrkoop D. Opioids for low back pain. *BMJ*. 2015;350:g6380. doi:10.1136/BMJ.G6380.

Du H, Ghassemi MM, Feng M. The effects of deep network topology on mortality prediction. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE; 2016:2602-2605. doi:10.1109/EMBC.2016.7591263.

Edlund MJ, Martin BC, Russo JE, Devries A, Braden JB, Sullivan MD. The Role of Opioid Prescription in Incident Opioid Abuse and Dependence Among Individuals with Chronic Non-cancer Pain. *Clin J Pain*. November 2013:1. doi:10.1097/AJP.0000000000000021.

Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016;316(22):2402. doi:10.1001/jama.2016.17216.

Haffajee RL, Mello MM. Drug Companies' Liability for the Opioid Epidemic. *N Engl J Med*. 2017;377(24):2301-2305. doi:10.1056/NEJMp1710756.

Jamei M, Nisnevich A, Wetchler E, Sudat S, Liu E. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. doi:10.1371/journal.pone.0181173.

Nelson LS, Juurlink DN, Perrone J. Addressing the Opioid Epidemic. *JAMA*. 2015;314(14):1453-1454. doi:10.1001/jama.2015.12397.

Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. $\mathtt {Deepr}$: A Convolutional Net for Medical Records. *IEEE J Biomed Heal Informatics*. 2017;21(1):22-30. doi:10.1109/JBHI.2016.2633963.

O'Connell C, Azad TD, Mittal V, et al. Preoperative depression, lumbar fusion, and opioid use: an assessment of postoperative prescription, quality, and economic outcomes. *Neurosurg Focus*. 2018;44(1):E5. doi:10.3171/2017.10.FOCUS17563.

Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. https://arxiv.org/pdf/1711.05225.pdf. Accessed June 2, 2018.

Rudd RA, Seth P, David F, Scholl L. Increases in Drug and Opioid-Involved Overdose Deaths - United States, 2010-2015. *MMWR Morb Mortal Wkly Rep*. 2016;65(5051):1445-1452. doi:10.15585/mmwr.mm655051e1.

Zhang W, Li R, Deng H, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage*. 2015;108:214-224. doi:10.1016/J.NEUROIMAGE.2014.12.061.

10 Appendix

Here we present selected hyperparameter tuning results for both datasets. For full hyperparameter tuning results, see our Github.

Supplemental Table 1: Hyperparameter tuning (Encodings Dataset)

Learning Rate	Dropout Rate	Batch Size	Nodes (N)	Layers	Training Loss	Dev Loss	Training Accuracy	Dev Accuracy
0.001	0.45	8192	250	1	0.461	0.586	0.782	0.718
0.001	0.45	8192	250	2	0.451	0.604	0.79	0.707
0.001	0.45	8192	250	3	0.454	0.597	0.793	0.713
0.001	0.45	8192	300	1	0.449	0.602	0.787	0.709
0.001	0.45	8192	300	2	0.441	0.61	0.796	0.708
0.001	0.45	8192	300	3	0.442	0.602	0.8	0.712
0.001	0.45	8192	400	1	0.427	0.619	0.801	0.708
0.001	0.45	8192	400	2	0.418	0.622	0.808	0.707
0.001	0.45	8192	400	3	0.421	0.619	0.807	0.707
0.0001	0.5	8192	250	3	0.556	0.558	0.727	0.726
0.001	0.5	8192	250	3	0.531	0.561	0.752	0.724
0.01	0.5	8192	250	3	0.586	0.588	0.72	0.718
1.00E-05	0.5	8192	250	3	0.594	0.592	0.707	0.711
0.001	0.5	1024	250	3	0.503	0.568	0.766	0.721
0.001	0.5	128	250	3	0.512	0.567	0.758	0.725
0.001	0.5	2048	250	3	0.499	0.568	0.771	0.719
0.001	0.5	256	250	3	0.507	0.568	0.761	0.721
0.001	0.5	4096	250	3	0.505	0.567	0.764	0.724
0.001	0.5	512	250	3	0.502	0.566	0.763	0.72
0.001	0.5	8192	250	3	0.529	0.561	0.75	0.725
0.001	0.3	8192	1000	0	0.361	0.659	0.85	0.684
0.001	0.3	8192	1000	1	0.168	1.123	0.931	0.681
0.001	0.3	8192	1000	2	0.223	1.454	0.883	0.663
0.001	0.3	8192	2000	0	0.304	0.695	0.879	0.684
0.001	0.3	8192	2000	1	0.113	1.333	0.955	0.668
0.001	0.3	8192	2000	2	0.058	1.142	0.984	0.682
0.001	0.3	8192	250	5	0.288	0.824	0.881	0.683
0.001	0.3	8192	250	6	0.303	0.779	0.876	0.685
0.001	0.3	8192	250	7	0.324	0.773	0.867	0.679
0.001	0.3	8192	3000	0	0.257	0.73	0.906	0.677
0.001	0.3	8192	3000	1	0.084	1.37	0.972	0.671
0.001	0.3	8192	3000	2	0.045	1.211	0.991	0.673
0.001	0.3	8192	500	0	0.406	0.637	0.822	0.687
0.001	0.3	8192	500	1	0.254	0.876	0.892	0.691
0.001	0.3	8192	500	2	0.196	0.976	0.918	0.682
0.001	0.3	8192	500	5	0.179	0.962	0.951	0.637
0.001	0.3	8192	500	6	0.167	0.962	0.953	0.647
0.001	0.3	8192	500	7	0.155	0.967	0.951	0.661
0.001	0.5	8192	1000	0	0.388	0.627	0.834	0.698
0.001	0.5	8192	1000	1	0.176	1.07	0.923	0.685
0.001	0.5	8192	1000	2	0.171	1.315	0.917	0.681
0.001	0.5	8192	2000	0	0.32	0.671	0.871	0.688
0.001	0.5	8192	2000	1	0.07	1.207	0.982	0.677
0.001	0.5	8192	2000	2	0.153	1.375	0.929	0.685
0.001	0.5	8192	300	6	0.5	0.632	0.802	0.668

Supplemental Table 2: Learning rate selection (Principal Components Dataset)

Epochs	Learning rate	Dropout rate	L2 Regularization	Batch size	Nodes	Layers	Training loss	Dev loss	Training	Dev accuracy
100	1.00E-02	0	0	8192	500	3	8.814	8.771	45%	46%
100	1.00E-03	0	0	8192	500	3	0.215	1.745	91%	63%
100	1.00E-04	0	0	8192	500	3	0.404	0.677	82%	68%
100	1.00E-05	0	0	8192	500	3	0.546	0.554	73%	73%
100	1.00E-06	0	0	8192	500	3	0.562	0.561	72%	72%
100	1.00E-07	0	0	8192	500	3	0.663	0.662	64%	64%

Supplemental Table 3: Network architecture selection (Principal Components Dataset)

Nodes	**Layers**	**Training Accuracy**	**Dev Accuracy**
100	1	75%	72%
100	2	77%	70%
100	3	80%	69%
200	1	78%	70%
200	2	82%	68%
200	3	85%	64%
300	1	80%	68%
300	2	89%	66%
300	3	97%	66%
400	1	83%	68%
400	2	89%	66%
400	3	97%	65%
500	1	85%	68%
500	2	92%	67%
500	3	98%	66%
600	1	84%	68%
600	2	92%	67%
600	3	98%	65%