DEEP ARTIFICIAL NEURAL NETWORK MODELS OF NEURAL ENCODING IN VISION AND NEUROSTIMULATION

by

ELIJAH DOUGLAS CHRISTENSEN

BS, University of Washington, 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Neuroscience Program

2020

This thesis for the Doctor of Philosophy degree by

Elijah Douglas Christensen

has been approved for the

Neuroscience Program

by

Gidon Felsen, Chair

Joel Zylberberg, Advisor

Zachary Kilpatrick

Alon Poleg-Polsky

Angie Ribera

Date: 14-August 2020

Christensen, Elijah Douglas (BS, Neuroscience Program)

Deep Artificial Neural Network Models of Neural Encoding in Vision and Neurostimulation

Thesis directed by Assistant Professor Joel Zylberberg

**ABSTRACT**

There is a significant need for more effective treatments for neurological and psychiatric diseases. Implantable neurostimulators are increasingly used as new therapeutic options for these diseases. This work will discuss our approach to mitigate current limitations in two types of implantable neurostimulators: Deep brain stimulation in treating Parkinson’s disease and cortical prosthetics in vision.

Deep brain stimulators (DBS) are typically configured to deliver therapeutic stimulation constantly, which can produce unavoidable side-effects and needlessly drains power. Modulating stimulation adaptively, or closed-loop stimulation, could mitigate these issues but requires methods to accurately read physiologically relevant brain states, preferably using only the already implanted electrodes. Another class of implanted neurostimulators, cortical prosthetics, rely on accurate predictions of neural activity in the targeted brain area for arbitrary stimuli. Current models used to predict neural activity in primary visual cortex only achieve 35% predictability overall and predictability declines in subsequent areas of visual processing.

With a focus on DBS and cortical prosthetics, we highlight how deep artificial neural network (ANN) models can be leveraged as a tool in neuroscience for studying neural encoding and decoding. We apply an ANN model trained using supervised learning to decode sleep state continuously from “spectral fingerprints” contained in local field potential activity of DBS electrodes. Furthermore, we show that deep convolutional neural networks can be used to make more accurate predictions of cortical neural encoding of visual stimuli in both early (primary visual cortex) and late (inferior temporal cortex) stages of visual processing.

The form and content of this abstract are approved. I recommend its publication.

Approved: Joel Zylberberg

**ACKNOWLEDGEMENTS**

The following work presented in this dissertation would not have been possible without support and advice from many other people in my life. Thank you, Joel, for your direction, guidance, and advice. I’m forever grateful for your patient mentorship and support throughout my graduate research. I could not have made it through graduate school without the unwavering support of my wife, Heidi; you believed in me when I stopped believing in myself. Special thanks to Alon for thoughtful discussion and direction and to Gidon for providing a place to call home.

The contributions in Chapter 3 was published in Journal of Sleep Research by Christensen et al. (2019). EC, JZ, JAT and AA were responsible for conception and design of the study. AA acquired the data. EC and JZ conceived the model and developed the code for its application. EC, JZ and JAT analyzed and interpreted the data. EC and JAT drafted the manuscript. EC, AA, JZ and JAT critically revised the manuscript.

The contributions in Chapter 5 was published in Journal of Vision by Kindel, Christensen, and Zylberberg (2019). WK and JZ conceived the original project. EC and WK analyzed and interpreted the data. WK and EC drafted the manuscript. EC, WK, and JZ critically revised the manuscript.

I would also like to thank my support by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program as well as by the Canada First Research Excellence Fund (CFREF) and York University.

TABLE OF CONTENTS

CHAPTER

1. [INTRODUCTION 1](#_Toc99053931)

[Deep Brain Stimulation 1](#_Toc99053932)

[Neural Interfaces 2](#_Toc99053933)

[Bottom-up neural encoding models 3](#_Toc99053934)

[Top-down neural encoding models 4](#_Toc99053935)

[Summary 6](#_Toc99053936)

1. [MACHINE LEARNING AND COMPUTATIONAL NEUROSCIENCE 11](#_Toc99053937)

[Artificial Neural Networks as a Model of Neural Computation 11](#_Toc99053938)

[Artificial Units 11](#_Toc99053939)

[Layers 12](#_Toc99053940)

[Model archetypes 12](#_Toc99053941)

[Architectures 14](#_Toc99053942)

[Fully connected 15](#_Toc99053943)

[Convolutional 15](#_Toc99053944)

[Loss Function 16](#_Toc99053945)

[Learning Rules 17](#_Toc99053946)

[Forward Pass 17](#_Toc99053947)

[Backpropagation 18](#_Toc99053948)

[Optimizers 18](#_Toc99053949)

[Summary 20](#_Toc99053950)

1. [PREDICTING SLEEP STATES IN HUMAN PARKINSON’S DISEASE PATIENTS 22](#_Toc99053951)

[Introduction 22](#_Toc99053952)

[Materials and Methods 23](#_Toc99053953)

[Patient Demographics 23](#_Toc99053954)

[Signal processing and local field potentials 23](#_Toc99053955)

[Video‐PSG scoring 24](#_Toc99053956)

[Model description 25](#_Toc99053957)

[Hyperparameter optimization 25](#_Toc99053958)

[Results 26](#_Toc99053959)

[Model performance and validation 26](#_Toc99053960)

[Discussion 27](#_Toc99053961)

1. [REFERENCES 34](#_Toc99053962)

CHAPTER I

INTRODUCTION

Neurological and psychiatric disease represent a significant societal burden in both advanced and developing countries (Collins et al., 2011) and there is a significant need for more effective treatments. Recent advances in brain stimulation and recording technology have enabled development of long-desired treatment options for many of these diseases in the form of implantable devices that directly stimulate populations of neurons. Deep brain stimulation (DBS) is one of these implantable devices utilized to mitigate disease symptoms. Patients with DBS receive electrical pulses via electrodes implanted in their brain. DBS has become an established therapy for movement disorders (Parkinson’s Disease (PD) and essential tremor) (Perlmutter and Mink, 2006) as well as epilepsy and psychiatric diseases (Holtzheimer and Mayberg, 2011). Another group of implantable neurostimulators are neural interfaces, such as cortical prostheses, which aim to restore sight in patients with congenital or acquired blindness. The body of work presented here makes progress on two unsolved challenges limiting advances in implantable neurostimulators, namely DBS state detection and more accurate cortical encoding of visual stimuli.

Deep Brain Stimulation

DBS uses a surgically implanted stimulator to apply electrical pulses directly to the brain to mitigate symptoms of neurologic and psychiatric diseases. Historically, drugs have been the primary method of treating these diseases, but DBS has emerged as a promising alternative for patients who do not respond to pharmacotherapy. Parkinson’s disease (PD) was among the first FDA approved uses of DBS for mitigating the disease’s motor symptoms. When employed for treating PD, current best practice for DBS therapy uses constant stimulation even though its therapeutic benefits to motor symptoms are needed most when the patient is awake to suppress resting tremor or bradykinesia in movement initiation. Current implanted stimulators are used this way because they have no way to detect when stimulation is not needed, such as when the patient is asleep or when lower levels of stimulation are needed to correct resting tremor. This strategy of constant stimulation, or open-loop stimulation, is less power efficient and comes with side effects such as impaired cognition, speech, gait, and balance (Hariz et al., 2008). However, activating DBS stimulation only when necessary requires a robust method for discerning whether or not the patient's brain needs stimulation. For example, a closed-loop DBS system would read out the patient’s brain state and only deliver electrical pulses during periods when the patient is awake (Figure 1.1). Closed-loop DBS is more power efficient and would have less collateral side effects by only stimulating when necessary.

Neural Interfaces

Cortical prosthetics (Fig 1.2) are a form of neural interface used to restore sight in blind patients (Lorach et al., 2013). These implantable neurostimulators bypass lost or damaged neurons by stimulating the damaged neuron targets the same way the original neurons otherwise would. Cortical prostheses must reproduce the neural activity patterns that would typically be relayed naturally by neurons of the thalamic lateral geniculate nucleus (LGN) and retina when directly stimulating visual cortex. Neural encoding, our understanding of how neurons reformat and represent visual stimuli, is key to this goal of properly restoring sight. The ultimate test of our knowledge of neural encoding is to predict neural responses to stimuli. Unfortunately, current models of neural encoding still struggle to accurately predict neuron responses to natural image stimuli. Subsequent sections will review two distinct approaches to developing models capable of predicting cortical responses to visual stimuli: bottom-up encoding models and top-down encoding models.

Bottom-up neural encoding models

Bottom-up encoding models use experimentally derived properties to explain responses in later visual areas. Neurons in early stages of visual processing can be characterized by their response to very specific local features in an image. A visual processing neuron’s receptive field (RF) is useful for depicting the properties of an image that modulate the neurons activity. RF’s are typically represented in models by linear filters applied at the first stage of processing. The inner product (i.e. dot product) between the filter and corresponding image region predict a given neurons response to that image. The linear RF model was insufficient for predicting several non-linear properties of retinal ganglion cell (RGC) responses to white noise and even worse for more complex stimuli. Subsequent Linear-Nonlinear-Poisson (LNP) (Paninski et al., 2004) models were better predictors of RGC spike rates in responses to white noise image stimuli. LNP combines a linear spatial filter with a single static non-linearity. The LNP model predicts neural responses well for white noise stimuli but does not generalize well when used to predict responses to natural image stimuli. Generalized Linear Models (GLM) (Pillow et al., 2008) improve prediction accuracy by accounting for interactions between RGC’s.

Top-down neural encoding models

As opposed to bottom-up neural encoding models, top-down models attempt to explain neural encoding as a result of optimizing an overarching goal. The genome likely has insufficient capacity for specifying every neuronal connection (synapse) (Zador, 2019) so what mechanisms ensure that neurons are connected correctly? This has recently been referred to as the “brain wiring problem” (Hassan and Hiesinger, 2015). We’ll be looking specifically at how synaptic wiring is determined in the visual cortex. Before the eyes even open, molecular interactions and spontaneous activity of RGC’s guide development of the initial “coarse” connectivity between RGC’s in the eye, to the neurons of the lateral geniculate nucleus in thalamus (LGN) and on to the primary visual cortex (V1) (Del Rio and Feller, 2006; Katz and Shatz, 1996). After this retinotopic map is established, synaptic connectivity continues refinement but requires environmental stimuli (Pietro Berkes et al., 2011). Identifying this “unifying principle” that guides stimulus-dependent refinement of connectivity would help explain the structure of visual representations in V1 and beyond.

Sparse coding

Shortly after the discovery of simple and complex cells (Hubel and Wiesel, 1959), Horace Barlow proposed efficient coding (Barlow, 1961) as an explanation for the computations performed by neural circuits in sensory cortex. The efficient coding hypothesis posits that the overarching goal of sensory processing is to reduce the high information redundancy in stimuli from the physical environment. This view was strengthened by findings that the Gabor-like receptive fields of simple cells are an optimal basis set for natural scenes when optimizing for 1) representation sparsity and 2) image reconstruction (D. Field, 1987; Olshausen and D. J. Field, 1996). Due to the highly metabolic nature of neurons, sparse coding was proposed because of its metabolic and information efficient properties (Levy and Baxter, 1996). Sparse coding models were particularly influential after successfully predicting aspects of neural computations in retina (Atick and Redlich, 1992), thalamus (Dan et al., 1996) and V1 (Olshausen and D. J. Field, 1996).

Optimizing for efficient coding would predict information redundancy should decrease as it is processed and relayed by successive visual areas. Information redundancy decreases when the same information can be carried by fewer neurons, which occurs as visual information propagates from photoreceptors to RGCs and from retinal ganglion cells to the LGN in the thalamus (Figure 1.3). Instead of information redundancy decreasing, as would be predicted by efficient coding, anatomical evidence seems to indicate that information redundancy in primary visual cortex is likely higher than it is prior areas of visual processing (Barlow, 2001; Felleman and Van Essen, 1991). Furthermore, despite some modest successes at explaining the complex response properties of V2 (the next visual area after V1)

(Lee et al., 2008; Olshausen et al., 2001) subsequent findings (Pietro Berkes et al., 2009; Willmore et al., 2011) have shown that visual areas beyond V2 cannot be explained by the efficient coding hypothesis. Efficient coding alone as an objective is not sufficient for explaining response properties of neurons in higher level visual areas like V4 and inferior temporal cortex (IT).

Goal-directed convolutional neural networks

Barlow, when reflecting later on his original idea makes a prescient statement, perhaps without knowing it: “We now need to step back and take a more global view of the brain’s **task** in order to see what lies behind the importance of recognizing redundancy” (Barlow, 2001). Neural networks which optimize behaviorally relevant tasks (Yamins et al., 2014) have shown state of the art performance at predicting neuronal activity across the ventral visual stream.

Summary

Chapter 2 provides an introduction to artificial neural networks (ANN), their similarities and differences to biological neurons, and the machine learning techniques used to train them which serves as a foundation for the technical chapters that follow.

In Chapter 3 we demonstrate ANN models as a tool for decoding sleep state in real-time using only the signals available from intracranial DBS electrodes implanted in the basal ganglia of PD patients. Importantly, this model generalizes decoding to patients never seen by the model and may allow new ways to leverage implantable stimulators for therapeutic benefit.

Chapter 4 explores the effects using composite loss functions (recognize and visualize) during training on both learned representations and task performance. This work was motivated by the observation that visual processing areas are reactivated during visualization tasks indicating their dual role in visual processing and regenerating stimuli.

Finally, Chapter 5 demonstrates the utility of neural network models and machine learning techniques as a way to explain response properties of individual neurons. We use a convolutional neural network (CNN) to achieve performance comparable to state of the art at predicting activity of individual neurons evoked by natural image stimuli in macaque V1. Furthermore, we use this model generatively to explain response properties of cells outside of Hubel and Wiesel’s simple- or complex-cell designations.

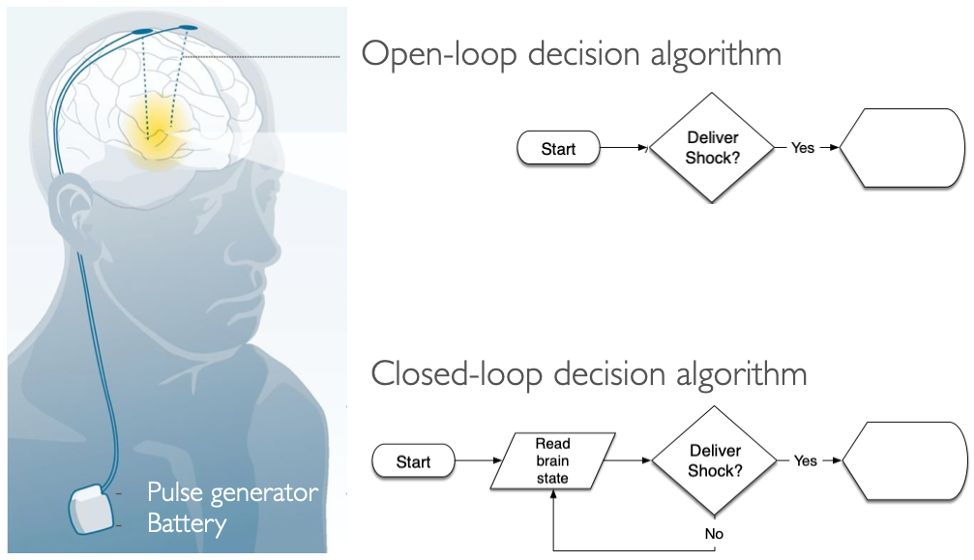


Figure 1.1 Current DBS systems use an open-loop design, where stimulation is delivered constantly without regard to the patient’s state. Closed-loop DBS system would read out the patient’s brain state to modulate stimulation intensity accordingly based on if the patient is awake, stationary, or moving to relieve symptoms of Parkinson’s Disease

A picture containing sitting, small, table, cake

Description automatically generated

Figure 1.2 Cortical Prosthetics. (A) Visual processing spans multiple brain regions, starting in the retina with retinal ganglion cells eventually progressing through primary visual cortex (V1) and inferior temporal cortex (IT). (B) Building a “camera-to-brain-translator” hinges on our ability to convert images or video into their equivalent cortical representations. We attempt to build better cortical encoding models using artificial neural networks, predicting neuron responses to images in V1 (Chapter 5) and IT (Chapter 4).

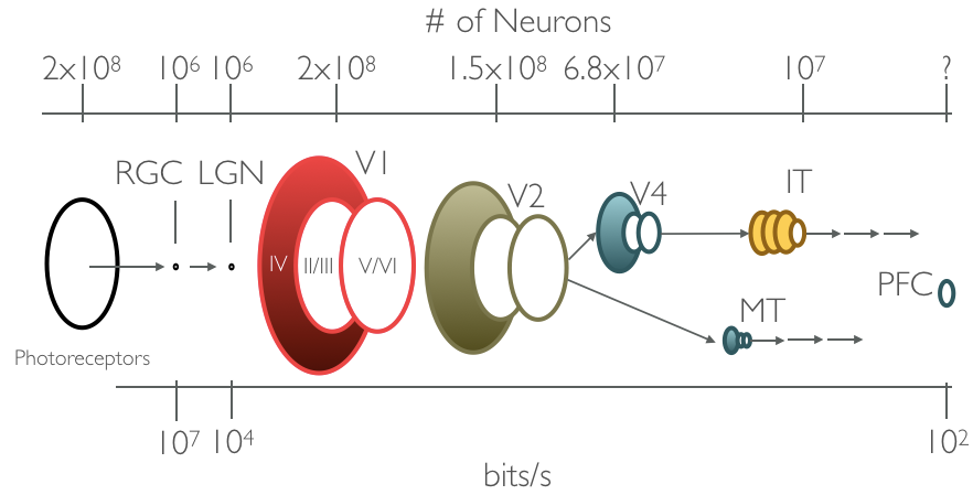


Figure 1.3 Channel capacity, the number of neurons carrying visual information, significantly varies along the ventral visual stream. Top axis denotes the estimated number of neurons in each area of visual processing in the ventral stream (Felleman and Van essen 1998). Bottom axis shows estimates of information through in bits/s at each area. Conveying the same information with reduced channel capacity is an example of reducing information redundancy.

CHAPTER II

MACHINE LEARNING AND COMPUTATIONAL NEUROSCIENCE

Artificial Neural Networks as a Model of Neural Computation

Despite the importance of computers for conducting machine learning and computational neuroscience research both fields had origins long before contemporary transistor computers. In 1943, inspired by the “all-or-none” nature of neural activity, Warren McCulloch (neuroscientist) and Walter Pitts (logician) formalized a simple mathematical definition of a neuron (McCulloch and Pitts, 1943). McCulloch and Pitts neurons became the fundamental unit of artificial neural networks (ANN). These artificial neurons, often referred to as (artificial) units, reproduce several key properties of real neurons (Fig 2.1). Biological neurons receive input from many other neurons via connections (e.g. synapses) to its dendrites. These synaptic inputs are summated at the soma where the net dendritic input increases or decreases the neurons membrane potential (Fig 2.1A). If the net dendritic input shifts the membrane potential beyond a certain threshold (e.g. the threshold potential) the neuron will fire action potentials.

Artificial Units

Artificial units (Fig 2.1B) are the basic building block of artificial neural networks. Each artificial unit receives input represented as a sequence of inputs and each input has a corresponding synaptic weight . In the artificial unit loosely represents a neurons membrane potential by adding net dendritic input to a scalar bias term () which is meant to represent the unit’s intrinsic excitability. Finally, the threshold non-linear response of biological neurons is captured by passing through an activation function () which gives the unit activation which is meant to loosely analogize neuronal firing rate.

Layers

Just as the brain is comprised of more than one neuron, most models make use of many artificial units. Similar to the functional organization of the neocortex, artificial neural networks (ANN) group individual units together in groups typically referred to as layers (Fig 2.1C). The artificial units within a layer collectively operate on a shared input and the layer’s output consists the collective activations of its constituent artificial units. ANN layers in a model between the inputs (x) and final outputs (y) are often referred to as “hidden” layers. The layers of an ANN are often considered analogous to a population of neurons in regions of the brain which perform similar functions. For instance, primary visual cortex (V1) contains a population of neurons which receive visual inputs from the retina (relayed by LGN). As a population of neurons, V1 processes this visual input and this processed visual information is then relayed to area V2 for subsequent processing and so on.

Model archetypes

Deep artificial neural network models typically have multiple layers stacked one after the other, such that the outputs of one layer become the inputs for the subsequent layer. Deep ANN models are often constructed for a specific purpose, or to perform a specific task. Models are often categorized based on purported task and the structure of the inputs it uses to accomplish this task. For instance, many computer vision researchers train models which, given an image, categorize the object in the image. The work presented in this thesis makes use of three distinct types of neural network models: classifiers, regressors and autoencoders. We will cover these model archetypes briefly in the following sections.

Classifiers

Classifiers are a class of models that attempt to predict the best category that describes the input from a discrete number of categories. For example, a classic machine learning exercise has been to train a model to predict the category of an object depicted in an image. MNIST, Fashion-MNIST(Xiao et al., 2017), CIFAR10/100(Krizhevsky and Hinton, 2009; Krizhevsky et al., n.d.) and ImageNet are examples of large labeled image datasets that have been historically popular for evaluating a model’s classification performance. Classifiers are not specific image tasks and can be used on any discrete labeling task. For instance, in Chapter 3 we trained an ANN classifier to predict behavioral sleep state in human PD patients based on features from local field potential spectral decompositions.

Regressors

Regressors use their inputs and attempt to predict a continuous value purportedly derived from the input. Recently, neural network models have been used as functional models of the visual system. These models use images to predict neuronal firing rates observed in animals after viewing the same image and they have been used to successfully for predicting stimulus evoked activity in retina (McIntosh et al., 2016) and Inferior Temporal cortex (IT) (Yamins et al., 2014). We successfully utilized a convolutional neural network regressor model to predict firing responses for populations of neurons in macaque primary visual cortex (V1) which is the subject of Chapter 5.

Autoencoders

Autoencoders are a special class of models which attempt to predict their inputs. This is a trivial task if each of the intermediate hidden layers have similar dimensionality as the input and output; the model can simply learn to copy the input into the output. Instead, these models are more often configured to have far fewer dimensions in their hidden layers. In this configuration the only way to successfully perform the task is to exploit information redundancy in the input to compress the input while retaining as much information as possible. We use an autoencoder in Chapter 4 to better capture the fact that the brain uses its representations for both recognition but also generatively in visualization.

Architectures

Training an ANN model using machine learning typically requires three components. These components are 1) the model’s layer architecture, 2) objective or loss function, and 3) the models learning rules. The layer architecture of a model explicitly specifies how the artificial units, organized in layers, are connected from input to output. There are a wide variety of layer types to choose from when constructing a deep ANN but for the sake of brevity only descriptions of layer architectures used in this work will be provided.

Fully connected

Fully-connected layers are the simplest and oldest of layer architectures. In all-to-all layers, every input is connected to every unit in the layer. We can describe this ANN layer mathematically by vectorizing the previous equation wherein inputs and output firing rates are represented as vectors () instead of scalars ():

Hyperbolic tangent (tanh), sigmoid, or Rectified Linear Units (ReLU) are often used as activation functions () but other more complex ones have also been introduced.

Convolutional

Convolutional layers have many parallels to (and were directly inspired by) the organization of the mammalian visual system. The early layers of visual processing are organized spatially, areas of field of view near each other are encoded near each other in RGCs, LGN, and V1 with nearly one-to-one correspondence. This matching topographic map is often referred to as retinotopic organization. Deeper layers than V1 retain some of this retinotopy but progressively pool these features representing larger and larger receptive field areas. Convolutional layers convolve a series of spatial filters across their 2D inputs to output “feature maps” of patches in the image that match the filter. They typically operate on images that have been separated into three distinct channels (RGB) and normalized as a surrogate for processing steps in retina and thalamus (Dan et al., 1996). However, this is not a defining characteristic, networks are also often trained on grayscale or color images that have not been preprocessed at all.

Loss Function

Loss or cost functions ( ) are mathematical definitions of the goal of the learning system. The loss function is used to calculate a scalar metric quantifying the models’ task performance as a function of their output. Loss functions can take any form mathematically, but typically differentiable loss functions are preferred for more straightforward optimization. Reconstruction error (sum of squared pixel errors) has traditionally been used for training models which attempt to generate a particular image. As an example of one loss function we can express the sum squared pixel loss between a model’s output image () and the target reference () as:

The target reference () is sometimes referred to as the teaching signal. In supervised learning the teaching signal is supplied to train the network the right answer for each particular batch of training examples.

Loss functions do not have to depend on a particular dataset or task. For instance, sparse coding models use activation sparseness and reconstruction error as their loss function to learn sparse representations. When minimized over images of natural scenes they learn to represent images using features that resemble localized receptive fields of simple cells in the primary visual cortex (Olshausen and D. J. Field, 1996; Zylberberg et al., 2011)

Learning Rules

Once a model’s architecture and loss function are specified “training it” is simply optimizing the parameters of each layer to improve its loss. The first algorithm for defining a method for iteratively updating the ANN model parameters to minimize loss was developed by Rumelhart (Rumelhart et al., 1986) and is still commonly used for training ANNs. We use this algorithm for training our models and it involves a simple 2 step process:

1. Forward pass: Use a batch of x input values to calculate the predicted outputs ()
2. Backpropagation: Use prediction error to update weights and biases

To illustrate this process, we will derive it for a simple 2-layer ANN. For simplicity, we change notation when describing deep ANN with multiple layers such that variable and function subscripts denote the variable or function’s corresponding layer NOT matrix or vector dimensions. For instance, we define the output activations at layer L in a model comprised of sequentially stacked all-to-all layers as:

Forward Pass

First, we pass a batch of training example inputs (x) through the model to get a batch of outputs (). Given our simple feedforward layer defined above, the full equation for the models output is given by:

For simplicity, we will combine all trainable parameters in this model into a variable

Our loss function ( ) defines how to evaluate the model’s performance as a function of the model’s predicted and target values. The target value is also sometimes referred to as the teaching signal, as it is used to teach the model the correct output for a given input. For this example, we’ll use sum-squared-error:

Backpropagation

To derive the gradient of the loss function with respect to the model parameters () we take a partial derivative of the loss function with respect to the model parameters:

Optimizers

Once we know the gradient of each weight with respect to the loss, we simply need to adjust the weights of the model in the direction specified by the weight gradient. Continually descending the gradient of the loss function should result in reaching a minimum of the loss for performing the model’s task but may not be the global minimum.

Stochastic Gradient Descent

Stochastic Gradient Descent (SGD) is the simplest and oldest optimization algorithm. Model parameters , are iteratively updated by subtracting the parameter gradient scaled by a learning rate according to the following equation

SGD optimization with momentum

Standard SGD works well if the surface of the loss function is smooth but has difficulty navigating “ravines”

(the1986, n.d.) which are common near local minima in optimization problems. Momentum (Qian, 1999) is a solution to this issue wherein a fraction of the previous update is added to the current weight update. This modification helps SGD accelerate in the relevant descent direction.

Picture a boulder accumulating speed as it rolls down a hill. Progressively increasing for dimensions gradient direction stays the same and reduces weight updates if the gradient direction changes.

ADAM optimization

Adaptive Moment Estimation (ADAM) computes adaptive learning rates for each parameter. If momentum is a ball rolling down a hill, ADAM optimization is a heavy ball with friction. It accomplishes this by computing decaying averages of past and past squared gradients ().

Summary

The purpose of this chapter is not to exhaustively cover the field of machine learning but instead to serve as a brief primer of concepts and terms you will encounter in subsequent chapters. Chapter 2 uses an ANN classifier comprised of fully-connected layers to predict sleep states from LFP spectral decompositions. Chapter 3 utilizes a convolutional autoencoder/classifier hybrid model to test hypotheses about computational objectives employed in primate ventral stream visual representations. Finally, Chapter 4 uses a convolutional neural network (CNN) to directly regress neuronal activity in macaque primary visual cortex.

Hopefully, you can appreciate the similarities between artificial neural networks and the biological neural networks that inspired them. If nothing else, remember that using machine learning to train ANN models hinges on three components:

1) Model architecture

2) Loss function

3) Learning rules

All three components influence both transient and final model performance.

A picture containing electronics, black, photo, monitor

Description automatically generated

Figure 2.1 (A) Connected neurons in the brain are the substrate of neural computation in brains. The connections between each neuron varies according to the strength of the synapse. Net dendritic input from other neurons is summated at the soma of the neuron. If the neuron is sufficiently depolarized at the soma it will fire an action potential down its axon. (B) Connected neurons are modeled in artificial neural networks by individually weighting each input and summing them capture net input, z, to the artificial unit. This net input is passed through a non-linear activation function, g, such as linear rectification or a sigmoid to capture the all-or-nothing behavior of biological neurons. (C) Individual units are grouped into layers. In a fully-connected layer, weight of each input, x, for each unit, h, is captured in the weight matrix W1. These unit activations form the inputs to subsequent layers, y, which have their own set of weights, W2.

CHAPTER III

PREDICTING SLEEP STATES IN HUMAN PARKINSON’S DISEASE PATIENTS[[1]](#footnote-1)

Introduction

Sleep is crucial to the regulation of physiological and cognitive functions in humans, and when disordered greatly diminishes quality of life (Giuditta et al., 1995; Pace-Schott and Hobson, 2002) and adversely affects nervous system repair (Brager et al., 2016; Lucke-Wold et al., 2015). Parkinson's disease (PD) is a neurodegenerative disorder that exhibits a high degree of comorbidity with a wide range of sleep disorders (De Cock et al., 2011; Tekriwal et al., 2017). The diagnosis and treatment of PD primarily focus on the overt motor symptoms (Postuma et al., 2015). However, there is increasing interest in understanding the impact of non‐motor symptoms, such as sleep dysfunction, on overall disease burden (Chaudhuri et al., 2006), and in identifying treatments for these symptoms. With the onset of motor fluctuations or breakthrough tremor despite optimal medical management, subthalamic nucleus (STN) deep brain stimulation (DBS) surgery has become the reference standard for treating the motor symptoms of advanced PD (Bronstein et al., 2011; Hamani et al., 2004). Interestingly, several studies have found that STN‐DBS can improve sleep in PD (Arnulf et al., 2000; De Cock et al., 2011; Iranzo et al., 2002). In our previous work, using local field potentials (LFPs) recorded from DBS electrodes implanted in STN for the treatment of PD, we identified unique spectral patterns within STN oscillatory activity that correlated with distinct sleep cycles, a finding that might offer insight into sleep dysregulation (Thompson et al., 2017). One extension of this work was to determine whether LFP information recorded from the STN could be used in real time to objectively identify sleep cycles for targeted therapy using DBS. In other words, the sleep benefit derived from STN stimulation could potentially be optimized using an adaptive stimulation algorithm that is aimed at specific sleep stages. In this study, we demonstrate the use of a feedforward artificial neural network that predicts sleep stage from LFP recordings, within the STN, with high precision.

Materials and Methods

Patient Demographics

This study was approved by the Institutional Review Board of the University of Minnesota, where the surgical and recording procedures were performed. All consenting study subjects (n = 9) carried a diagnosis of idiopathic PD (Figure 3.1a). Subjects were unilaterally implanted in the STN with a quadripolar DBS electrode (model #3389: Medtronic Inc., Fridley, MN), per routine surgical protocol (Abosch et al., 2012). Experimental details for the recording setup have been previously published (Thompson et al., 2017). Basic characterization of these data was previously reported in Thompson et al. (2017).

Signal processing and local field potentials

Signal processing of the raw STN LFP signals was previously described in Thompson et al. (2017). Briefly, after preprocessing, the four LFP channels (0, 1, 2 and 3; one recording from each of the four electrical contacts of the implant) were converted into three bipolar derivations (LFP01, LFP12 and LFP23) by sequentially referencing them. Power spectral density (PSD) was estimated using a fast Fourier transform from a 2‐s‐long sliding window (Hamming) with 1‐s overlap. The final time‐evolving spectra had 15 s time and 0.5 Hz frequency resolution. For each subject, LFP data selected for further analysis were based on the location of the DBS electrode contact within the STN and this was verified by the following: (a) intraoperative microelectrode recordings that identified cells with firing characteristics consistent with STN neurons; (b) anti‐Parkinsonian benefit and side‐effects of macrostimulation; (c) preoperative stereotactic T1‐ and T2‐weighted images merged to a postoperative MRI demonstrating the position of the DBS electrode within the borders of STN; (d) the use of Framelink (Medtronic Corp.) software to analyze DBS position on the postoperative MRI; and (e) evaluation of the efficacy of post-programming stimulation for contralateral motor symptoms for each subject (Ince et al., 2010). Selection of which contact(s) to use for study recordings was based on the STN contact (s) associated with peak beta‐spectrum activity as this feature correlates with the optimal programming contact(s) for the treatment of contralateral motor symptoms (Ince et al., 2010). These criteria were used to ensure that the selected contact was most reliably in the same relative anatomical location across patients to permit generalizability of the model.

Video‐PSG scoring

The polysomnographic electrode montage used was the following: F3–C3, P3–O1, F4–C4 and P4–O2, EOGL–A2, EOGR–A1, and chin EMG (Iber et al., 2007). Sleep stages were determined by analysis of 30‐s epochs of the PSG, by a sleep neurologist, with each epoch classified as Awake or as belonging to one of the following sleep stages: rapid eye movement (REM), or the non‐REM (NREM) stages of N1, N2 or N3.

Model description

We trained a feedforward artificial neural network (ANN) with a single hidden layer (Figure 2b) to prospectively identify whether a given 30‐s epoch of STN‐LFP recording took place during one of three possible states: REM, NREM or Awake. Inputs to the model were eight separate frequency band power bins, averaged over 30 s: delta (0–3 Hz), theta (3–7 Hz), alpha (7–13 Hz), low beta (13–20 Hz), high beta (20–30 Hz), and low gamma (30–90 Hz), high gamma (90–200) and high frequency oscillations (200–350). Each frequency range input feature was normalized independently by subtracting the mean and scaling by the variance of feature. The ANN output is a probability that the measured epoch occurs during one of the three possible states. Optimal ANN architecture was chosen based on the hyperparameter optimization detailed below. The ANN model utilizes a single hidden layer to encode the normalized spectral power bands within 32 features by calculating weighted sums of the input frequency power and scaling them by a non‐linear function. Weighted linear combinations of these 32 features are then used by the network to compute sleep state probabilities with application of a softmax non‐linearity.

Hyperparameter optimization

The architecture of the ANN model we describe was determined by evaluating classification accuracy across the spectrum of network hyperparameters. We combinatorically varied the non‐linearity of each unit (Sigmoid, ReLu and Tanh), the number of units in the hidden layer(s) (16, 32 or 64) and the number of hidden layers (1 or 2). Randomly initialized models in replicates of five were each trained and tested on a random 80:20 partition of all data. In general, we observed that more complex models with a larger number of total units and multilayer networks produced minor increases in classification accuracy, but these performance variations were not statistically significant. We opted to use 32 units in a single hidden layer with the biologically‐inspired rectified linear units (ReLu; (Hahnloser et al., 2000) ) as the non-linearity. We chose this configuration because it achieved classification accuracy on a par with the best‐performing model with 10‐fold fewer parameters to minimize overfitting training data.

Results

Model performance and validation

We evaluated the ANN model's sleep stage classification performance and its ability to generalize new predictions under two conditions. Performance was evaluated using accuracy and Cohen's κ. Chance accuracy was calculated as originally described

(Cohen, 1960).

First, we tested the model's ability to predict sleep stages on novel examples from patients included in the training set. We pooled 80% of each patient's 30‐s STN‐LFP recording epochs across all nine patients to train the model. The remaining 20% of the withheld epochs were used to evaluate the model's performance on novel examples from familiar patients. The train‐test fractions (80:20) were sampled randomly for each patient and performance was averaged in replicates of five to prevent sampling bias. The model was able to correctly predict sleep stage from STN‐LFP epochs with a mean accuracy of 91% (Figure 3.3a).

Training a model from scratch for each new patient is often intractable. Therefore, the model's ability to perform well on never‐ seen subjects demonstrates its sensitivity to the salient spectral features of sleep across individual variations. To test this level of generalization, the model was trained on all epochs from eight of the nine patients. Subsequently, model performance was evaluated on all epochs from the kept‐out patient. Thus, nine different models were trained, each with a specific patient withheld from its training data. As above, model performance was quantified using accuracy and Cohen's κ (Figure 3.3b). Across all models, mean classification accuracy of 91% was observed. Finally, because the number of epochs of each observed sleep state varies between patients in the dataset, we produced confusion matrices for the test patient of each model and show representative examples from patients with significantly imbalanced sampling as well as a summary matrix averaged across all models (Figure 3.3c). This demonstrates that the model's error rate varies as a function of sleep‐ stage representation, with less frequent stages showing a higher error rate (see Table 3.1).

Discussion

In this report, we demonstrate the novel use of an optimized ANN to predict sleep stage from 30‐s epochs of LFP recorded from the STN of PD subjects. Based on results from hyperparameter optimization, we used a network architecture of a single hidden layer containing 32 artificial neurons with ReLu non‐linearities (Figure 3.2b). We evaluated the model's ability to generalize to new patients by using a LOGO (leave‐one‐group‐out) strategy for cross‐validation and attained mean classification accuracy of 91% averaged across all patients.

The ability of this ANN model to accurately predict sleep stages based on STN‐LFP data recorded from novel PD patients is a critical improvement over our previously published effort to generate a predictive model. In our prior work, we used a support vector machine (SVM) model that performed well when tested on novel epochs derived from the familiar patient used to train the model but failed to generalize to novel subjects (Thompson et al., 2017). For simplification of model development, the different NREM stages (i.e. NREM 1–3) were aggregated into a single class. However, future development will focus on classification of the non‐REM substages, as they represent distinct states and underlie unique sleep processes. Our current study is the first to use direct intracranial recordings from human basal ganglia to classify and match unseen PSG‐labelled electrophysiological signals. Although the overall accuracy of the model for all sleep stages combined was well above chance (91%), performance on REM sleep stages was lower than the average performance (77%). Decreased performance for REM could be a result of the lower representation across subjects (see Table 3.1), or it may reflect the challenge in identifying the REM state from PSG in this patient population. This model can be implemented in forthcoming improved DBS neurostimulators to detect sleep stage solely from features of STN‐recorded LFP, enabling the implementation of closed‐loop stimulation strategies for treating sleep dysregulation in PD patients. This would serve a crucial unmet need in this patient population (Chaudhuri et al., 2006), as there are currently no effective treatments with a low side‐effect burden(Arnulf et al., 2000). Although DBS is an established therapy for the treatment of motor symptoms of Parkinson's disease, the effect of DBS on the sleep disturbances of Parkin- son's disease has not yet been fully characterized, and the mechanism(s) underlying the improvements reported in sleep quality, efficiency and duration remains to be elucidated(Sharma et al., 2018).

Our model's ability to correctly predict sleep stage in novel subjects may imply the existence of a universal LFP spectrum sleep signature within STN. In our investigations to date, this STN localized spectral signature appears conserved across patient demographics, robust to variances in implantation location, and detectable from the aggregate activity of several thousands of neurons. In future work, we intend to characterize this spectral signature space using generative ANN models of LFP oscillations recorded from within the STN. This effort will extend our understanding of the relationship between sleep dynamics and oscillating field potentials in the basal ganglia.

A picture containing screenshot

Description automatically generated

Figure 3.1 (a) Demographic data and sleep stage characteristics for Parkinson's disease (PD) subjects participating in this study (n = 9). Percent improvement in PD reflects the change in the Unified Parkinson's Disease Rating Scale (UPDRS) motor scale before and after DBS surgery. (b) Hypnograms from four representative subjects in this study, indicative of common sleep architecture deficits reported for individuals with PD. (c) Distribution of frequency band power contribution to sleep stage for all subjects. AWM, awake with movement; AWOM, awake without movement; REM, rapid eye movement

A close up of text on a white background

Description automatically generated

Figure 3.2 (a) Representative spectrogram of a local field potential (LFP) recording acquired over the course of one full night's sleep from a deep brain stimulation (DBS) electrode implanted into the subthalamic nucleus (STN). A PSG‐ informed hypnogram assessed by a sleep expert is aligned with the LFP recordings (red line; AWM, awake with movement; AWOM, awake without movement; REM, rapid eye movement; N1–3, non‐rapid eye movement stages 1–3). (b) Schematic representation of the feedforward classifier used to predict sleep stage from 30‐s labelled LFP epochs. The model is composed of an input layer (LFP frequency power bands), a hidden layer and an output layer (predicted sleep stage). (c) Comparison of hypnogram assessed by a sleep expert (top; black) and ANN‐ predicted hypnogram (bottom; red) from patient 1 with mean classification accuracy of 87%

A close up of a map

Description automatically generated

Figure 3.3 (a) In the “hybrid” strategy a random 80% of each patient's local field potential (LFP) recordings were pooled to train the model. Model accuracy and Cohen's κ were evaluated on the withheld 20% from each patient. This analysis was replicated in four other random 80:20 splits to control sampling bias. Cohen's κ magnitude guidelines derived from Fleiss & Cohen (1973). (b) A leave‐one‐group‐out (LOGO) cross‐validation strategy was used to test generalizability to unseen patients. Each data point represents a model trained with a specific patient excluded from its training data. Model accuracy and Cohen's κ were evaluated on data from the kept‐out patient. (c) Confusion matrices of representative models trained using the LOGO cross‐validation strategy. The first two confusion matrices represent individual subjects and the final confusion matrix depicts the fraction of epochs with specific class labels for all subjects. REM, rapid eye movement; NREM, non‐rapid eye movement

Table 3.1 Summary for all subjects of the epoch representation and model accuracy for each of the following sleep stages: Awake, rapid eye movement (REM) and an aggregate of the non‐rapid eye movement (NREM) substages (N1, N2 and N3)

**A screenshot of a cell phone

Description automatically generated**

REFERENCES

Abosch, A., Lanctin, D., Onaran, I., Eberly, L., Spaniol, M., Ince, N.F., 2012. Long-term recordings of local field potentials from implanted deep brain stimulation electrodes. Neurosurgery 71, 804–814. doi:10.1227/NEU.0b013e3182676b91

Arnulf, I., Bejjani, B.P., Garma, L., Bonnet, A.M., Houeto, J.L., Damier, P., Derenne, J.P., Agid, Y., 2000. Improvement of sleep architecture in PD with subthalamic nucleus stimulation. Neurology 55, 1732–1734. doi:10.1212/wnl.55.11.1732

Atick, J.J., Redlich, A.N., 1992. What Does the Retina Know about Natural Scenes? Neural Computation 4, 196–210. doi:10.1162/neco.1992.4.2.196

Barlow, H., 2001. Redundancy reduction revisited. Network 12, 241–253.

Barlow, H.B., 1961. The coding of sensory messages. Current Problems in Animal Behavior 331–360.

Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research 13, 281–305.

Brager, A.J., Yang, T., Ehlen, J.C., Simon, R.P., Meller, R., Paul, K.N., 2016. Sleep Is Critical for Remote Preconditioning-Induced Neuroprotection. Sleep 39, 2033–2040. doi:10.5665/sleep.6238

Bronstein, J.M., Tagliati, M., Alterman, R.L., Lozano, A.M., Volkmann, J., Stefani, A., Horak, F.B., Okun, M.S., Foote, K.D., Krack, P., Pahwa, R., Henderson, J.M., Hariz, M.I., Bakay, R.A., Rezai, A., Marks, W.J., Moro, E., Vitek, J.L., Weaver, F.M., Gross, R.E., DeLong, M.R., 2011. Deep brain stimulation for Parkinson disease: an expert consensus and review of key issues., in:. Presented at the Archives of neurology, American Medical Association, pp. 165–165. doi:10.1001/archneurol.2010.260

Cadena, S.A., Denfield, G.H., Walker, E.Y., Gatys, L.A., Tolias, A.S., Bethge, M., Ecker, A.S., 2018. Deep convolutional models improve predictions of macaque V1 responses to natural images. bioRxiv 201764. doi:10.1101/201764

Cadieu, C.F., Hong, H., Yamins, D.L.K., Pinto, N., Ardila, D., Solomon, E.A., Majaj, N.J., DiCarlo, J.J., 2014. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLoS Comput. Biol. 10, e1003963. doi:10.1371/journal.pcbi.1003963

Carandini, M., Heeger, D.J., 2011. Normalization as a canonical neural computation. Nat Rev Neurosci 13, 51–62. doi:10.1038/nrn3136

Chaudhuri, K.R., Healy, D.G., Schapira, A.H.V., National Institute for Clinical Excellence, 2006. Non-motor symptoms of Parkinson's disease: diagnosis and management. Lancet Neurol 5, 235–245. doi:10.1016/S1474-4422(06)70373-8

Chen, Y., Crawford, J.D., 2019. Allocentric representations for target memory and reaching in human cortex. Annals of the New York Academy of Sciences 46, 774. doi:10.1111/nyas.14261

Coen-Cagli, R., Kohn, A., Schwartz, O., 2015. Flexible gating of contextual influences in natural vision. Nat. Neurosci. 18, 1648–1655. doi:10.1038/nn.4128

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales Educational and Psychological Measurement, vol. 20.

Collins, P.Y., Patel, V., Joestl, S.S., March, D., Insel, T.R., Daar, A.S., Scientific Advisory Board and the Executive Committee of the Grand Challenges on Global Mental Health, Anderson, W., Dhansay, M.A., Phillips, A., Shurin, S., Walport, M., Ewart, W., Savill, S.J., Bordin, I.A., Costello, E.J., Durkin, M., Fairburn, C., Glass, R.I., Hall, W., Huang, Y., Hyman, S.E., Jamison, K., Kaaya, S., Kapur, S., Kleinman, A., Ogunniyi, A., Otero-Ojeda, A., Poo, M.-M., Ravindranath, V., Sahakian, B.J., Saxena, S., Singer, P.A., Stein, D.J., 2011. Grand challenges in global mental health. Nature 475, 27–30. doi:10.1038/475027a

Dan, Y., Atick, J.J., Reid, R.C., 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. Journal of Neuroscience 16, 3351–3362.

David, S.V., Vinje, W.E., Gallant, J.L., 2004. Natural Stimulus Statistics Alter the Receptive Field Structure of V1 Neurons. Journal of Neuroscience 24, 6991–7006. doi:10.1523/JNEUROSCI.1422-04.2004

De Cock, V.C., Debs, R., Oudiette, D., Leu, S., Radji, F., Tiberge, M., Yu, H., Bayard, S., Roze, E., Vidailhet, M., Dauvilliers, Y., Rascol, O., Arnulf, I., 2011. The improvement of movement and speech during rapid eye movement sleep behaviour disorder in multiple system atrophy. Brain 134, 856–862. doi:10.1093/brain/awq379

Del Rio, T., Feller, M.B., 2006. Early retinal activity and visual circuit development. Neuron 52, 221–222. doi:10.1016/j.neuron.2006.10.001

DiCarlo, J.J., Zoccolan, D., Rust, N.C., 2012. How does the brain solve visual object recognition? Neuron 73, 415–434. doi:10.1016/j.neuron.2012.01.010

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.

Field, D., 1987. Relations between the statistics of natural images and the response properties of cortical cells. Journal of Optical Society of America.

Freud, E., Plaut, D.C., Behrmann, M., 2016. “What” Is Happening in the Dorsal Visual Pathway. Trends in Cognitive Sciences 20, 773–784. doi:10.1016/j.tics.2016.08.003

Geirhos, R., Temme, C.R.M., Rauber, J., Schuett, H.H., Bethge, M., Wichmann, F.A., 2018. Generalisation in humans and deep neural networks.

Giuditta, A., Ambrosini, M.V., Montagnese, P., Mandile, P., Cotugno, M., Grassi Zucconi, G., Vescia, S., 1995. The sequential hypothesis of the function of sleep. Behav. Brain Res. 69, 157–166. doi:10.1016/0166-4328(95)00012-i

Güçlü, U., van Gerven, M.A.J., 2015. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. J. Neurosci. 35, 10005–10014. doi:10.1523/JNEUROSCI.5023-14.2015

Hahnloser, R.H.R., Sarpeshkar, R., Mahowald, M.A., Douglas, R.J., Seung, H.S., 2000. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature 405, 947–951. doi:10.1038/35016072

Hamani, C., Saint-Cyr, J.A., Fraser, J., Kaplitt, M., Lozano, A.M., 2004. The subthalamic nucleus in the context of movement disorders. Brain 127, 4–20. doi:10.1093/brain/awh029

Hariz, M.I., Rehncrona, S., Quinn, N.P., Speelman, J.D., Wensing, C., Multicentre Advanced Parkinson's Disease Deep Brain Stimulation Group, 2008. Multicenter study on deep brain stimulation in Parkinson's disease: an independent assessment of reported adverse events at 4 years. Mov. Disord. 23, 416–421. doi:10.1002/mds.21888

Hassan, B.A., Hiesinger, P.R., 2015. Beyond Molecular Codes: Simple Rules to Wire Complex Brains. Cell 163, 285–291. doi:10.1016/j.cell.2015.09.031

Holtzheimer, P.E., Mayberg, H.S., 2011. Deep brain stimulation for psychiatric disorders. Annu. Rev. Neurosci. 34, 289–307. doi:10.1146/annurev-neuro-061010-113638

Hong, H., Yamins, D.L.K., Majaj, N.J., DiCarlo, J.J., 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. Nat. Neurosci. 19, 613–622. doi:10.1038/nn.4247

Hubel, D.H., Wiesel, T.N., 1959. Receptive fields of single neurones in the cat's striate cortex. The Journal of Physiology 148, 574–591.

Iber, C., Ancoli-Israel, S., Chesson, A., Quan, S., 2007. The AASM manual for the scoring of sleep and associates events: Rules, Terminology, and technical specifications. Westchester, IL: American Academy of Sleep Medicine.

Ince, N.F., Gupte, A., Wichmann, T., Ashe, J., Henry, T., Bebler, M., Eberly, L., Abosch, A., 2010. Selection of optimal programming contacts based on local field potential recordings from subthalamic nucleus in patients with Parkinson's disease. Neurosurgery 67, 390–397. doi:10.1227/01.NEU.0000372091.64824.63

Iranzo, A., Valldeoriola, F., Santamaría, J., Tolosa, E., Rumià, J., 2002. Sleep symptoms and polysomnographic architecture in advanced Parkinson's disease after chronic bilateral subthalamic stimulation. J Neurol Neurosurg Psychiatry 72, 661–664. doi:10.1136/jnnp.72.5.661

Katz, L.C., Shatz, C.J., 1996. Synaptic activity and the construction of cortical circuits. Science 274, 1133–1138. doi:10.1126/science.274.5290.1133

Krizhevsky, A., Hinton, G., 2009. Learning multiple layers of features from tiny images.

Krizhevsky, A., Nair, V., Hinton, G.E., n.d. CIFAR10.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in:. Presented at the Advances in Neural Information Processing Systems NIPS, pp. 1097–1105.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. doi:10.1038/nature14539

Lee, H., Ekanadham, C., Ng, A.Y., 2008. Sparse deep belief net model for visual area V2 873–880.

Lehky, S.R., Sejnowski, T.J., Desimone, R., 1992. Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. Journal of Neuroscience 12, 3568–3581. doi:10.1523/JNEUROSCI.12-09-03568.1992

Levy, W.B., Baxter, R.A., 1996. Energy efficient neural codes. Neural Computation 8, 531–543. doi:10.1162/neco.1996.8.3.531

Li, Q., Hu, J., Ding, J., Zheng, G., 2014. Fisher's method of combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations. Biostatistics 15, 284–295. doi:10.1093/biostatistics/kxt045

Lorach, H., Marre, O., Sahel, J.-A., Benosman, R., Picaud, S., 2013. Neural stimulation for visual rehabilitation: advances and challenges. J. Physiol. Paris 107, 421–431. doi:10.1016/j.jphysparis.2012.10.003

Lucke-Wold, B.P., Smith, K.E., Nguyen, L., Turner, R.C., Logsdon, A.F., Jackson, G.J., Huber, J.D., Rosen, C.L., Miller, D.B., 2015. Sleep disruption and the sequelae associated with traumatic brain injury. Neurosci Biobehav Rev 55, 68–77. doi:10.1016/j.neubiorev.2015.04.010

Mahendran, A., Vedaldi, A., 2015. Understanding deep image representations by inverting them, in:. Presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 5188–5196. doi:10.1109/CVPR.2015.7299155

Majaj, N.J., Hong, H., Solomon, E.A., DiCarlo, J.J., 2015. Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. J. Neurosci. 35, 13402–13418. doi:10.1523/JNEUROSCI.5181-14.2015

Marr, D., Poggio, T., 1976. From understanding computation to understanding neural circuitry. A.I. Memo 357

Mazurek, M., Kager, M., Van Hooser, S.D., 2014. Robust quantification of orientation selectivity and direction selectivity. Front Neural Circuits 8, 92. doi:10.3389/fncir.2014.00092

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity., Bulletin of mathematical biology.

McIntosh, L.T., Maheswaranathan, N., Nayebi, A., Ganguli, S., Baccus, S.A., 2016. Deep Learning Models of the Retinal Response to Natural Scenes. arXiv 29, 1369–1377.

Montijn, J.S., Meijer, G.T., Lansink, C.S., Pennartz, C.M.A., 2016. Population-Level Neural Codes Are Robust to Single-Neuron Variability from a Multidimensional Coding Perspective. Cell Rep 16, 2486–2498. doi:10.1016/j.celrep.2016.07.065

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A Toolbox for Representational Similarity Analysis. PLoS Comput. Biol. 10, e1003553. doi:10.1371/journal.pcbi.1003553

O'Craven, K.M., Kanwisher, N., 2006. Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. http://dx.doi.org/10.1162/08989290051137549 12, 1013–1023. doi:10.1162/08989290051137549

O'Craven, K.M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. J Cogn Neurosci 12, 1013–1023.

Olshausen, B.A., Field, D.J., 2005. How Close Are We to Understanding V1? Neural Computation 17, 1665–1699. doi:10.1162/0899766054026639

Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609. doi:10.1038/381607a0

Olshausen, B.A., Sallee, P., Lewicki, M., 2001. Learning sparse image codes using a wavelet pyramid architecture. papers.nips.cc

papers.nips.cc.

Ozbay, B.N., Losacco, J.T., Cormack, R., Weir, R., Bright, V.M., Gopinath, J.T., Restrepo, D., Gibson, E.A., 2015. Miniaturized fiber-coupled confocal fluorescence microscope with an electrowetting variable focus lens using no moving parts. Optics Letters 40, 2553–2556. doi:10.1364/OL.40.002553

Pace-Schott, E.F., Hobson, J.A., 2002. The neurobiology of sleep: genetics, cellular physiology and subcortical networks. Nat Rev Neurosci 3, 591–605. doi:10.1038/nrn895

Paninski, L., Pillow, J.W., Simoncelli, E.P., 2004. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural Computation 16, 2533–2561. doi:10.1162/0899766042321797

Perlmutter, J.S., Mink, J.W., 2006. Deep brain stimulation. Annu. Rev. Neurosci. 29, 229–257. doi:10.1146/annurev.neuro.29.051605.112824

Pietro Berkes, Ben White, Fiser, J., 2009. No evidence for active sparsification in the visual cortex 108–116.

Pietro Berkes, Orbán, G., Lengyel, M., Fiser, J., 2011. Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment. Science 331, 83–87. doi:10.1126/science.1195870

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., Simoncelli, E.P., 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454, 995–999. doi:10.1038/nature07140

Postuma, R.B., Berg, D., Stern, M., Poewe, W., Olanow, C.W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A.E., Halliday, G., Goetz, C.G., Gasser, T., Dubois, B., Chan, P., Bloem, B.R., Adler, C.H., Deuschl, G., 2015. MDS clinical diagnostic criteria for Parkinson's disease. Mov. Disord. 30, 1591–1601. doi:10.1002/mds.26424

Qian, N., 1999. On the momentum term in gradient descent learning algorithms. Neural Netw 12, 145–151. doi:10.1016/s0893-6080(98)00116-6

Richards, B.A., Lillicrap, T.P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R.P., de Berker, A., Ganguli, S., Gillon, C.J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G.W., Miller, K.D., Naud, R., Pack, C.C., Poirazi, P., Roelfsema, P., Sacramento, J., Saxe, A., Scellier, B., Schapiro, A.C., Senn, W., Wayne, G., Yamins, D., Zenke, F., Zylberberg, J., Therien, D., Kording, K.P., 2019. A deep learning framework for neuroscience. Nat. Neurosci. 22, 1761–1770. doi:10.1038/s41593-019-0520-2

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. doi:10.1038/323533a0

Schoppe, O., Harper, N.S., Willmore, B.D.B., King, A.J., Schnupp, J.W.H., 2016. Measuring the Performance of Neural Models. Front Comput Neurosci 10, 1929. doi:10.3389/fncom.2016.00010

Sereno, A.B., Lehky, S.R., 2011. Population coding of visual space: comparison of spatial representations in dorsal and ventral pathways. Front Comput Neurosci 4, 159. doi:10.3389/fncom.2010.00159

Sharma, V.D., Sengupta, S., Chitnis, S., Amara, A.W., 2018. Deep Brain Stimulation and Sleep-Wake Disturbances in Parkinson Disease: A Review. Front Neurol 9, 697. doi:10.3389/fneur.2018.00697

Simonyan, K., Zisserman, A., n.d. Very deep convolutional networks for large-scale image recognition. arxiv.org

arXiv:1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15, 1929–1958.

Stokes, M., Thompson, R., Cusack, R., Duncan, J., 2009. Top-down activation of shape-specific population codes in visual cortex during mental imagery. J. Neurosci. 29, 1565–1572. doi:10.1523/JNEUROSCI.4657-08.2009

Tekriwal, A., Kern, D.S., Tsai, J., Ince, N.F., Wu, J., Thompson, J.A., Abosch, A., 2017. REM sleep behaviour disorder: prodromal and mechanistic insights for Parkinson's disease. J Neurol Neurosurg Psychiatry 88, 445–451. doi:10.1136/jnnp-2016-314471

the, R.S.P.O.T.E.A.C.O., 1986, n.d. Two problems with back propagation and other steepest descent learning procedures for networks, ci.nii.ac.jp

.

Thompson, J.A., Tekriwal, A., Felsen, G., Ozturk, M., Telkes, I., Wu, J., Ince, N.F., Abosch, A., 2017. Sleep patterns in Parkinson’s disease: direct recordings from the subthalamic nucleus. J Neurol Neurosurg Psychiatry 89, jnnp–2017–316115–104. doi:10.1136/jnnp-2017-316115

Vintch, B., Movshon, J.A., Simoncelli, E.P., 2015. A Convolutional Subunit Model for Neuronal Responses in Macaque V1. J. Neurosci. 35, 14829–14841. doi:10.1523/JNEUROSCI.2815-13.2015

Willmore, B., Prenger, R.J., Wu, M.C.K., Gallant, J.L., 2008. The berkeley wavelet transform: a biologically inspired orthogonal wavelet transform. Neural Computation 20, 1537–1564. doi:10.1162/neco.2007.05-07-513

Willmore, B.D.B., Mazer, J.A., Gallant, J.L., 2011. Sparse coding in striate and extrastriate visual cortex. Journal of Neurophysiology 105, 2907–2919. doi:10.1152/jn.00594.2010

Xiao, H., Rasul, K., Vollgraf, R., 2017. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms.

Yamins, D.L.K., DiCarlo, J.J., 2016. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 19, 356–365. doi:10.1038/nn.4244

Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. U.S.A. 111, 8619–8624. doi:10.1073/pnas.1403112111

Zador, A.M., 2019. A Critique of Pure Learning: What Artificial Neural Networks can Learn from Animal Brains. bioRxiv 29, 582643. doi:10.1101/582643

Ziemba, C.M., Freeman, J., Movshon, J.A., Simoncelli, E.P., 2016. Selectivity and tolerance for visual texture in macaque V2. Proceedings of the National Academy of Sciences 113, E3140–E3149. doi:10.1073/pnas.1510847113

Zylberberg, J., DeWeese, M.R., 2013. Sparse Coding Models Can Exhibit Decreasing Sparseness while Learning Sparse Codes for Natural Images. PLoS Comput. Biol. 9, 1–10. doi:10.1371/journal.pcbi.1003182

Zylberberg, J., Murphy, J.T., DeWeese, M.R., 2011. A Sparse Coding Model with Synaptically Local Plasticity and Spiking Neurons Can Account for the Diverse Shapes of V1 Simple Cell Receptive Fields. PLoS Comput. Biol. 7, 1–12. doi:10.1371/journal.pcbi.1002250

1. This chapter was previously published in Christensen, Abosch, Thompson, and Zylberberg (2019)  *Journal of Sleep Research* 28:e12806 and are included with the permission of the copyright holder. [↑](#footnote-ref-1)