

Do biological constraints impair dendritic computation?

Ilenna Simone Jones^{1,*} and Konrad Paul Kording^{1,2}

¹*Department of Neuroscience, University of Pennsylvania*

²*Department Bioengineering, University of Pennsylvania*

**Corresponding author: Ilenna Jones, ilennaj@pennmedicine.upenn.edu*

August 12, 2021

Abstract

Computations on the dendritic trees of neurons have important constraints. Voltage dependent conductances in dendrites are not similar to arbitrary direct-current generation, they are the basis for dendritic nonlinearities and they do not allow converting positive currents into negative currents. While it has been speculated that the dendritic tree of a neuron can be seen as a multi-layer neural network and it has been shown that such an architecture could be computationally strong, we do not know if that computational strength is preserved under these biological constraints. Here we simulate models of dendritic computation with and without these constraints. We find that dendritic model performance on interesting machine learning tasks is not hurt by these constraints but may benefit from them. Our results suggest that single real dendritic trees may be able to learn a surprisingly broad range of tasks.

Introduction

The dominating point-neuron idea (McCulloch and Pitts, 1943) which underlies artificial neural networks (ANNs) (Lecun et al., 2015; Krizhevsky et al., 2012; Minsky and Papert, 1969) assumes that dendrites are linear integrators of synaptic inputs. However dendrites are highly nonlinear due to their voltage dependent ion channels (Antic et al., 2010). Detailed models of dendritic processing thus use multi-compartment modeling with differing levels of abstraction (Hines and Carnevale, 1997; Izhikevich, 2007) allowing for good models for phenomena such as direction sensitivity (Rall, 1964; Koch et al., 1982) and coincidence detection (Agmon-Snir et al., 1998; London and Häusser, 2005). However, we do not just want to ask if a realistic neuron can do one thing, but also to understand the general class of problems a neuron can solve. In this vein, the perceptron learning rule has been used to train a biophysical multi-compartment model to do binary classification tasks (Moldwin and Segev, 2019). We know that real neurons can solve interesting problems, e.g. the XOR task (Gidon et al., 2020), which requires at least two layers in an ANN. Indeed, recent computational work (David et al., 2019) has shown that a single dendritic tree can only be simulated with a many layered neural network of limited width. From these, it is clear that a dendrite can do more than linearly integrate, but we must also consider that there are important biological constraints. This raises the question of how the biological properties of dendrites affect their computational capabilities.

Owing to their rich nonlinear properties, real dendrites have been seen as equivalent to a multi-layer ANN (Poirazi et al., 2003b,a; Mel, 2016; Poirazi and Papoutsis, 2020). Specifically, it is generally seen as a multi-layer ANN that is constrained to have a tree structure where inputs are successively combined with one another on the way to the soma where the spike generator sits. Inspired by these ideas, our own past work modeled a single neuron as using a binary tree structure (Jones and Kording, 2021). These ANN neuron models utilize one important biological constraint, the tree structure of the dendritic tree, and test its impact on neuronal output and computation respectively. However, these studies only looked at a small set of biological constraints.

This lack of biological constraints results in models abstracted away from biological reality. Here we focus on three of these abstractions in our own tree-structured single neuron model (Jones and Kording, 2021) and identify ways it could be more biologically plausible. First, its nonlinear activation function is a Leaky Rectified Linear Unit (LReLU). Though LReLU may be a widely used activation function in many successful deep ANNs (Nair and Hinton, 2010; Krizhevsky et al., 2012; Xu et al., 2015; Maas et al., 2013), perhaps we can rely on voltage-dependent ion channel nonlinearities to identify a better activation function. Second,

its inputs are directly received by the model as values between 0 and 1. Perhaps we can make input more realistic by mapping this input through a conductance-based synaptic nonlinearity in the units of millivolts (Koch, 1999). Third, its weight parameters can be both positive and negative, allowing the activations traveling through the tree to flip signs. The weights in the binary tree structure of our model would be analogous to axial conductance between compartments (Hines and Carnevale, 1997), and so an improvement could be to constrain these weights to be non-negative (Rall, 1967). These are ways we can move a highly abstracted model closer toward biological realism. However, we suspect that imposing further biological constraints may yield differing performance results.

Here we use these suggested improvements as three biological constraints for our binary tree model: synaptic input nonlinearities, voltage-gated-ion-channel-derived dendritic nonlinearities, and non-negative axial conductance weights. We observe the impacts of introducing each constraint by training and testing the binary tree model on 7 different binary classification machine learning datasets. We find that the dendritic nonlinearity activation function introduced performs as well if not better than other typical activation functions. We find that synaptic input constraints generally do not limit or expand computation performance. Interestingly we see that the non-negative weight constraint limits performance in low multi-synaptic-input cases, but the combination of this non-negative weight constraint and the synaptic input constraint eliminates this limitation. Lastly, we find that the binary tree model with all three constraints performs as well or better than the model without these constraints. In conclusion, implementing these constraints fulfills the dual goal of creating a more biologically realistic binary tree neuron model and showing us that these combined constraints don't limit, and may actually improve, the model's computational capability. This suggests real dendritic trees may be able to perform a broad range of complex functions.

Background

Local dendritic dynamics

Neuroscience has a rich set of methods for the simulation of neurons, based on cable equations with an addition of a host of neuron-specific knowledge. The resulting simulations can generate the time-varying activities of a neuron in response to arbitrary stimuli. Here we use an approximation that is different in two ways. First, we only want to consider constant inputs. Second, we only want to ask if the voltage will keep going up and a spike will result and we are not interested in the dynamics of the spike itself. For those cases we will make a steady state approximation for currents going from the periphery towards the soma. This dendritic nonlinearity or activation function, which we call the NaCaK function (as it is

meant to capture sodium, calcium, and potassium currents), is but one of many potential approaches to representing the nonlinear local dendritic dynamics.

When we have a node in the neuron there will be currents coming in from the parent dendritic branches (which have currents I_1 and I_2), we will have local sodium (I_{Na}), calcium (I_{Ca}), and potassium (I_K) currents, we will have the current relating to charging the local capacitance (I_C), and currents going towards the soma (I_{out}),

$$I_{total} = I_1 + I_2 + I_{Na} + I_{Ca} + I_K + I_C - I_{out}. \quad (1)$$

For the maximal conductances ($g_{\bar{N}a}$, $g_{\bar{C}a}$, $g_{\bar{K}}$) and current-voltage (IV) curves of sodium (Hodgkin and Huxley, 1952), potassium (Doiron et al., 2001), and calcium (Miyasho et al., 2001) channels there exists an extensive literature. We used IV curves that were derived from biophysical simulations of each ion channel using BRIAN2 (Branco and Häusser, 2010), wherein the peak current was recorded for a range of voltage clamp settings. Given the current peaks for these ion channels occurred near instantaneously (within 1 millisecond) and we assume we are only dealing in local steady state dynamics. We then fit continuous functions to each of the IV curves, yielding the following equations,

$$f_{Na}(V) = \frac{0.0878(113.68\text{mV} - V)}{1\text{mV} + \exp(6.39\text{mV} - V/8.98)} \quad (2)$$

$$f_{Ca}(V) = \frac{0.129(69.62\text{mV} - V)}{1 + \exp(-4.40\text{mV} - V/4.25)} \quad (3)$$

$$f_K(V) = \frac{-2.23\text{mV}}{0.436\text{mV} + \exp(-0.132(V - 16.74\text{mV}))}. \quad (4)$$

We thus have a set of meaningful functions for local currents that form the NaCaK function.

The dynamics on the dendritic tree are complex and time-varying, and contain spiking dynamics. Instead of simulating such a complex model we will use a very simple approximation:

$$V \approx \frac{1}{K}(g_1 V_1 + g_2 V_2 + g_{\bar{N}a} f_{Na}(V_0) + g_{\bar{C}a} f_{Ca}(V_0) + g_{\bar{K}} f_K(V_0)). \quad (5)$$

We set $g_{\bar{N}a}$, $g_{\bar{C}a}$, and $g_{\bar{K}}$ to equal 1, and $\frac{1}{K} = 0.5$ in our model.

Our dendritic nonlinearity function, the NaCaK function, takes on the role of the activation function in more traditional ways of conceptualizing dendrites as neural networks (Poirazi et al., 2003a; Jones and Kording, 2021). Our derivation of this result of course made problematic assumptions about how to solve the relevant implicit equations. Another way of seeing the NaCaK function is simply a way of conceptually capturing the nonlinear effects induced by the three channel types.

Local synaptic dynamics approximation derivation

Just like in the case of the overall currents and local nonlinearities above, we will approximate the synapse properties, which are the input layer in our model. Using standard biophysical components, we can derive the steady state voltage by considering capacitive (I_C), membrane (I_R), excitatory (I_e), inhibitory (I_i), leak (I_{leak}), and "injected" (I_{inj}) currents,

$$I_{total} = I_C + I_R + I_e + I_i + I_{leak} + I_{inj}. \quad (6)$$

We can replace these current variables with their biophysical terms, while also ignoring membrane currents given the area of a synapse is small. For simplicity, both excitatory and inhibitory currents are included in the same synapse, so this formulation does not follow Dale's law. $I_{inj} = 0$ in this case as well,

$$I_{total} = C \frac{dV}{dt} + g_e(V - E_e) + g_i(V - E_i) + g_0(V - E_0). \quad (7)$$

The conductances g_e and g_i are functions of time and could be approximated from the slope of an I-V plot of peak currents of excitatory and inhibitory channels (Koch, 1999). The positive values of g_e and g_i will be calculated using a set of parameters α that characterize each synapse, can be trained and define the conductances:

$$g_p(x) = \exp(\alpha_{1p}x + \alpha_{2p}). \quad (8)$$

Where the input variable x , which is between 0 and 1, can be interpreted as probability of vesicular release. The parameter α_{1p} can be interpreted as an approximation of number of quantal release sites or vesicle size. The parameter α_{2p} can be interpreted as peak

conductance as a result of triggering post-synaptic receptors, which would be a function of however many post-synaptic receptors are present. This formulation is similar to that which is elaborated on in Koch (1999) (p. 91). However, it is different because it is an exponential expression, which we used to force the synaptic conductance to be positive and differentiable. Lastly, g_0 is a trainable parameter, which means there are 5 trainable parameters for each synapse ($\alpha_{1e}, \alpha_{2e}, \alpha_{1i}, \alpha_{2i}$, and g_0). These components in this formulation add more synaptic parameters (degrees of freedom) replacing what is typically assumed to be a scalar value.

The steady state convergence voltage where $\frac{dV}{dt} = 0$ can then be calculated:

$$V_\infty = \frac{g_e E_e + g_i E_i + g_0 E_0}{g_e + g_i + g_0}. \quad (9)$$

We set E_e , E_i and E_0 to be 50 mV, -70 mV and -65 mV to correspond to excitatory, inhibitory, and leak channels at the synapse.

We thus have a meaningful approximation of the effect of a synapse on the downstream dendrites.

Results

In this paper we modify the previously established ANN dendritic neuron model, called the k-tree (Jones and Kording, 2021), and add three more dendritic constraints to it in a dual effort of making it more biologically plausible and observing the impact of these constraints on model performance when they are introduced. We introduce NaCaK nonlinearity (Figure 1B, Figure S1), which is a linear combination of sodium, calcium, and potassium IV curves that map voltage input to current output. We introduce a synapse nonlinearity (Figure 1C), which is a steady state synaptic voltage response as a function of a continuous value that represents vesicular dynamics. This synapse nonlinearity output is in millivolts. Lastly, we introduce an axial conductance weight constraint (Figure 1D) where the weight parameters between each node is constrained to be non-negative since in neuron models conductance is a scalar value. As a control, we use a parameter-size-comparable 2-layer Fully Connected Neural Network (FCNN) dense network to see the impact of the sparse binary tree structure. We also have multi-synaptic-input repetition conditions to see the impact of repeated inputs to different dendritic sub-trees. (Figure 1A). Implementing these biological constraints and comparing their presence with their absence allows us to observe the computational impact of these constraints.

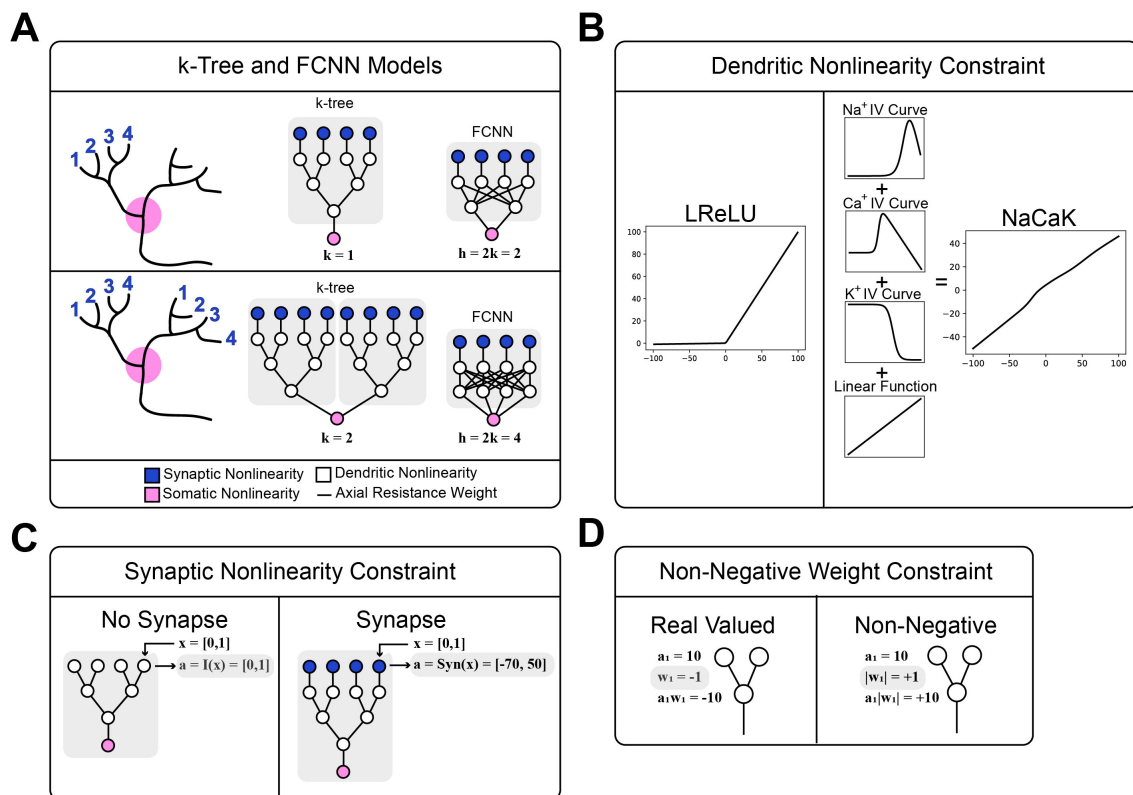


Figure 1: Binary k -tree Model and Fully Connected Neural Network (FCNN) Control Model Containing Three Biological Constraints. A) Blue nodes apply an optional synaptic nonlinearity to the input. White nodes apply an optional dendritic nonlinearity. Black lines connecting the nodes are weights that are optionally constrained to be non-negative. The value k determines how many sub-trees of identical input are repeated in the model as an analogy to repeated synaptic inputs on different dendritic trees. The value h corresponds to the number of nodes in the hidden layer of the FCNN. B) Comparison of LReLU activation function and NaCaK activation function. C) Comparison of model without synapse nonlinearity and with synapse nonlinearity. D) Comparison of model weights without and with the non-negative weight constraint.

NaCaK nonlinearity outperforms other nonlinearities in k -tree model

In densely connected ANNs, Rectified Linear Unit (ReLU) and LReLU have been accepted to improve model performances. However since we are using a sparsely connected model architecture, the k -tree, we are interested in seeing if the NaCaK function differentially impacts k -tree model performance. To do this we tested the k -tree performance on various datasets with the NaCaK activation function and compare this performance to the ReLU, LReLU, Sigmoid, and Swish activation functions. We find that the NaCaK nonlinearity outperforms ReLU, LReLU, and sigmoid activation functions in 5 out of 7 datasets in the k -tree model (Figure 2A-E, Figure S3A-E). Interestingly, the swish nonlinearity also performs

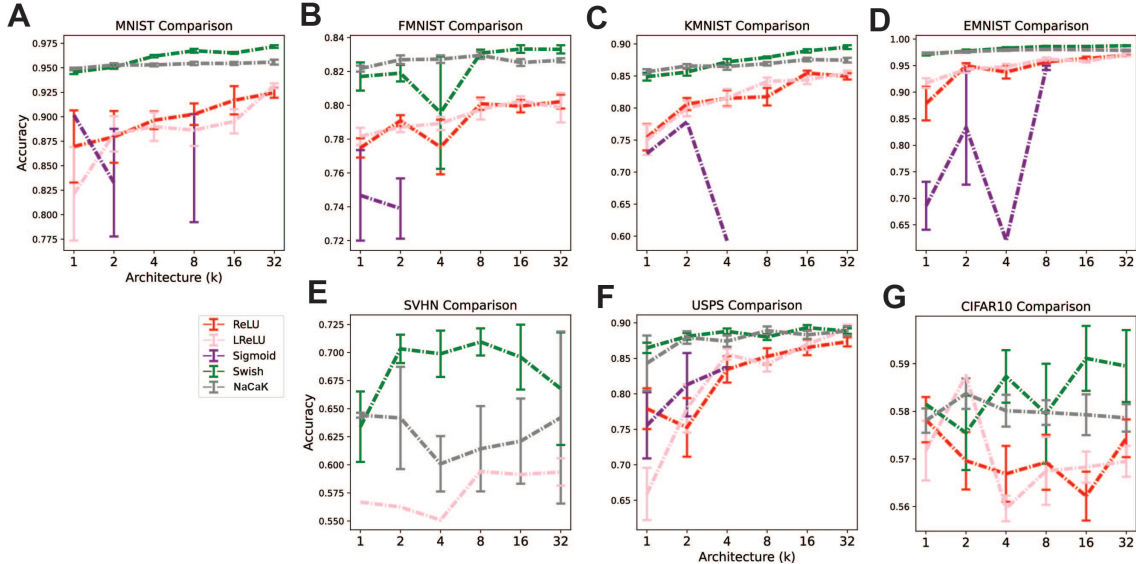


Figure 2: Comparing Nonlinear Activation Functions for the Dendrite Nodes in the k -tree For Each Dataset. Architecture ranges from low input multi-synaptic input repetitions ($k=1$) to high input repetitions ($k=32$) and mean accuracy of each model is displayed with standard error bars. Number of trials ranges from 1-10, and trial accuracy was omitted if the training was deemed to fail (accuracy threshold 0.55). Table 2 lists trial counts. Red: Rectified Linear Unit (ReLU), Pink: LReLU, Purple: Sigmoid, Green: Swish, Grey: NaCaK Nonlinearity

about as well or better than the NaCaK nonlinearity in the k -tree model. (Figure 2A-F, Figure S3A-F). For the FCNN model, ReLU and LReLU architectures performed better than all other nonlinearities in 4 out of 7 datasets (Figure S2A-D,G, Figure S4A-D,G). These results show that the NaCaK and swish functions paired with the binary tree architecture perform better than other nonlinearities.

Addition of synaptic input nonlinearity can improve performance of binary tree model

Without synapse nonlinearity mapping of the inputs, the k -tree model is receiving inputs that are not biologically realistic. We want to see if using the synapse nonlinearity constraint (Figure 1C), which nonlinearly maps the input to realistic millivolt units, impacts k -tree model computation. In order to show this, we implement a steady-state voltage conception of synaptic dynamics in order to map the input values to synaptic voltage output. Every input corresponds to a differing synaptic nonlinearity function. We allow the synaptic nonlinearities to have 5 different trainable parameters ($\alpha_{1e}, \alpha_{2e}, \alpha_{1i}, \alpha_{2i}, g_0$) instead of simply 1 trainable weight. We thus compare the densely connected FCNN and the binary tree models under the synapse or no-synapse conditions, all using the NaCaK nonlinearity, in order to observe

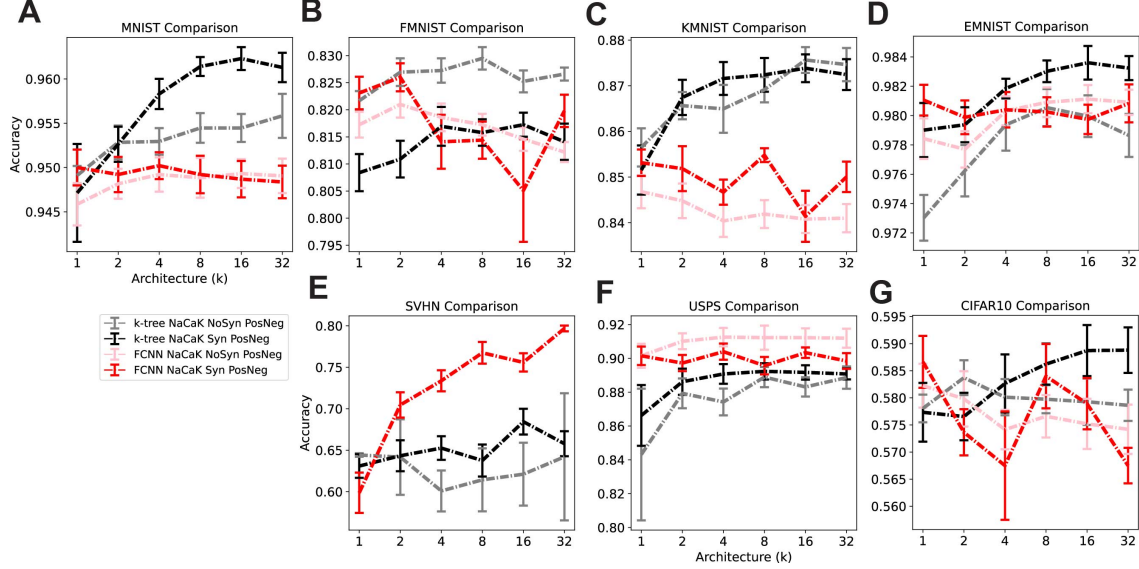


Figure 3: Comparing k -tree and FCNN models With or Without Synapse nonlinearity. Architecture ranges from low input multi-synaptic input repetitions ($k=1$) to high input repetitions ($k=32$) and mean accuracy of each model is displayed with standard error bars. Number of trials ranges from 1-10, and trial accuracy was omitted if the training was deemed to fail (accuracy threshold 0.55). Table 3 lists trial counts. Red: FCNN with Synapses, Pink: FCNN without Synapses, Black: k -tree with Synapses, Grey: k -tree without Synapses.

their impacts on computation.

Given that we look at 7 datasets, each of which belies a different result for how the presence of the synapse nonlinearity impact computation, it is difficult to draw a general conclusion from these results. Before considering the synapse condition, it is important to acknowledge that for the MNIST, FMNIST, KMNIST and EMNIST datasets, the binary tree outperformed the FCNN in the high multi-synaptic input repetition conditions and matched it in the low repetition conditions (Figure 3A-D, Figure S5A-D). We had assumed that the densely connected FCNN with a similar number of parameters as the sparsely connected k -tree would always have better performance. This result shows that this assumption is not the case. Also, if we are to simply compare the k -tree NaCaK conditions, then in MNIST and EMNIST datasets, the synapse nonlinearity architecture improves performance compared to the no-synapse architecture in the high-repetition conditions (Figure 3A,D, Figure S5A,D). The opposite may be seen in the FMNIST dataset (Figure 3B, Figure S5B), and it there is no difference in the KMNIST, SVHN, USPS, and CIFAR datasets (Figure 3C,E-G, Figure S5C,E-G). In the MNIST and KMNIST datasets, the k -tree performs better than the FCNN regardless of the presence of the synapse (Figure 3A,C, Figure S5A,C). In the FMNIST dataset, k -tree without synapse performs better than both FCNN conditions (Figure 3B, Figure S5B). From this variety of results, we can say that the combination of a binary tree

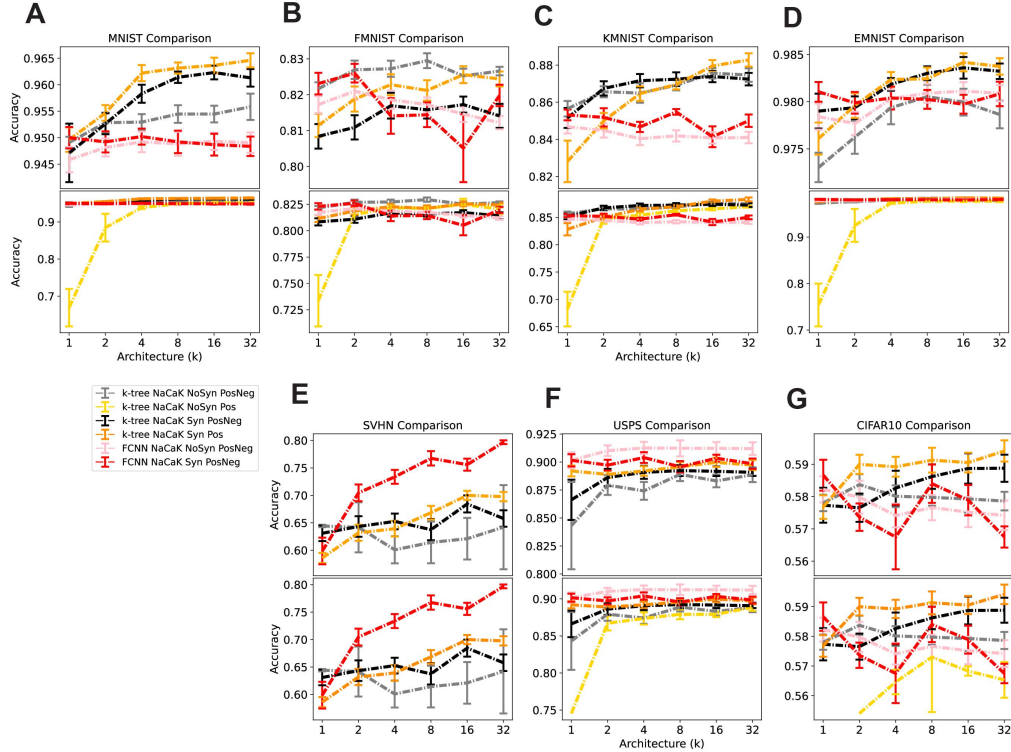


Figure 4: Comparing k -tree and FCNN models With or Without the Non-negative Weight Constraint. Architecture ranges from low input multi-synaptic input repetitions ($k=1$) to high input repetitions ($k=32$) and mean accuracy of each model is displayed with standard error bars. Number of trials ranges from 1-10, and trial accuracy was omitted if the training was deemed to fail (accuracy threshold 0.55). Table 4 lists trial counts. Orange: k -tree with Synapses and with Non-negative weights, Yellow: k -tree without Synapses and with Non-negative weights Red: FCNN with Synapses, Pink: FCNN without Synapses, Black: k -tree with Synapses, Grey: k -tree without Synapses. Top graphs in A-G omit the accuracies of the k -tree without synapses and with non-negative weights in order to resolve the lines with higher accuracies. Bottom graphs include this omitted set of accuracies.

constraint and a synaptic nonlinearity does not worsen model performance relative to all other conditions, and even has the potential to improve model performance above the supposed upper bound set by the FCNN model in higher-repetition cases.

A k -tree model with all biological constraints performs as well as or better than a model missing any one constraint

In the binary tree model, the weight parameters connecting each node with downstream nodes can be argued to be analogous to axial conductances that is often found in multi-compartment models of neurons (Figure 1D). Conductance is a scalar value, therefore its important to observe the impact of introducing the non-negative weight constraint in our specialized ANN

model. In order to observe this constraint’s impact on computation, we implement the k -tree with NaCaK nonlinearity architectures with or without synapse nonlinearities, and add the condition of constraining the model to have non-negative weights only. In Figure 4 we compare all previous with or without synapse conditions to these new conditions.

We thus show that in all datasets, the non-negative k -tree without synapse performs very poorly in the low-multi-synaptic input repetition regime, or does not perform above our training failure accuracy threshold (0.55) at all (Figure 4A-G: Bottom, Figure S5A-G). Importantly, this should be compared to the non-negative k -tree with synapses, which performs at a level comparable to other conditions (Figure 4, Figure S5). The non-negative k -tree with synapses performs at the level or better than of the positive and negative k -tree with synapses (Figure 4, Figure S5). In MNIST, KMNIST, and CIFAR10 conditions, the model with all constraints outperform FCNN with or without synapses (Figure 4A,C,G, Figure S5A,C,G) From this we can conclude that, using the k -tree NaCaK nonlinearity architecture, the combination of the presence of synapses and non-negative weights performs much better than the absence of synapses with non-negative weights. Overall, the most realistic k -tree model, with NaCaK nonlinearity, synapse nonlinearity and non-negative weights, performs at the level or better than a model missing any one of these constraints.

Experimental Procedures

Much of this section can be derived from Jones and Kording (2021), and has been modified for the purpose of this study.

Computational Tasks

Knowing that the output of a neuron is binary (presence or absence of an action potential), we chose to train our neuron model on a binary classification task. Using standard, high-dimensional, computer vision datasets, we used linear discriminant analysis (LDA) linear classifier to determine which 2 classes within each dataset were least linearly separable through training the LDA linear classifier and testing it on pairs of classes (Table 1). We used MNIST (Lecun et al., 1998), Fashion-MNIST (Xiao et al., 2017), EMNIST (Cohen et al., 2017), Kuzushiji-MNIST (Clanuwat et al., 2018), CIFAR-10 (Krizhevsky, 2009), Street View House Numbers (SVHN) (Goodfellow et al., 2014), and USPS (Hastie et al., 2001) datasets.

MNIST	FMNIST	EMNIST	KMNIST	CIFAR10	SVHN	USPS
0.8753 \pm 0.0120	0.6750 \pm 0.0108	0.5821 \pm 0.0180	0.6790 \pm 0.0164	0.5254 \pm 0.0069	0.5186 \pm 0.0102	0.8362 \pm 0.0306

Table 1: Linear Classifier Performance on Machine Learning Datasets

Controls

The control we use is a fully connected neural network (FCNN). The 2-layer FCNN is a comparable reference to see if k -tree performance meets or exceeds that of a densely connected network. The hidden layer of the FCNN is equal to twice the number of trees ($2k$) in the k -tree it is compared to and its output layer has 1 node.

Data Preprocessing

We used datasets from the torchvision (version 0.5.0) python package. We then padded the 28 by 28 resolution images with zeros so that they were 32 x 32, and flattened the images to 1-D vectors. We then split the shuffled training set into training and validation sets (for MNIST, the ratio was 1:5 so as to let the validation set size match the test set), Then we split the resultant shuffled training set and validation set into 10 independent subsets. Each subset was used for a different cross-validation trial.

Model Architecture

Using Pytorch (version 1.6.0), we designed the k -tree model architecture to be a feed forward neural network with sparse binary-tree connections. The weight matrices were implemented as simulated sparse tensors using the SparseLinear package created by rain-neuromorphics (updated August 18, 2020). Each node received 2 inputs and produces 1 output. To account for the sparsification, we altered the initialization of the weight matrices: we used standard “Kaiming normal” initialization with the gain of 1/density of the weight matrices. All nonzero weights were trained in model optimization. Lastly, the leak hyperparameter for LReLU was set to 0.01.

Model Training

The model, inputs, and labels were loaded onto a Nvidia GeForce 1080 GPU using CUDA version 10.1. The batch size was 256. Early stopping was used such that after 60 epochs where no decrease in the validation loss is observed, training is stopped. We allowed the parameters of sparse linear layers of the k -tree and the synapse nonlinearity parameters to

train. We did not allow the NaCaK conductance parameters ($g_{\bar{N}a}$, $g_{\bar{C}a}$ and $g_{\bar{K}}$) to train. In all conditions, loss was calculated using mean binary cross entropy loss. In the k -tree condition with the synapse layer implemented, we also added a mean hinge loss when activations at any node were outside of the range of -70 and 50. We used the python library Hyperopt, which is based on the Tree-structured Parzen Estimator algorithm, in order to optimize the learning rates used for our Adam optimizer for each dataset and constraint condition. Each train-test loop was run for 10 trials with a different training subset each trial and the same test set every trial. Trial averages and standard errors were then calculated, and p-values were determined using two-tailed student's t-test (Figure S3-S5).

Ion Channel Model Parameters

The NaCaK function was built out of the IV curves of sodium (Hodgkin and Huxley, 1952), calcium (Miyasho et al., 2001), and potassium (Doiron et al., 2001) channels. The IV curves were derived from biophysical simulations of each ion channel using BRIAN2 (Branco and Häusser, 2010), wherein the peak current was recorded for a range of voltage clamp settings. The parameters for these IV curves are depicted below.

The sodium channel (Hodgkin and Huxley, 1952) model had maximum conductance $g_{\bar{N}a} = 120\text{mS}$ and a reversal potential $E_{N_a} = 115\text{mV}$.

$$\begin{aligned}
 I_m &= g_{N_a} m^3 h (V - E_{N_a}) & (10) \\
 \alpha_m &= \frac{1}{\exp(\frac{25\text{mV}-V}{10\text{mV}})} \\
 \beta_m &= 4 \exp(\frac{-V}{18\text{mV}}) \\
 \alpha_h &= 0.07 \exp(\frac{-V}{20\text{mV}}) \\
 \beta_h &= \frac{1}{1 + \exp(\frac{30\text{mV}-V}{10\text{mV}})}.
 \end{aligned}$$

The calcium channel (Miyasho et al., 2001) model had maximum conductance $g_{\bar{C}a} = 100\text{mS}$ and a reversal potential $E_{C_a} = 135\text{mV}$.

$$\begin{aligned}
I_m &= g_{Ca} m h (V - E_{Ca}) \\
\alpha_m &= \frac{2.6}{1 + \exp(\frac{V+21\text{mV}}{-8\text{mV}})} \\
\beta_m &= \frac{0.18}{1 + \exp(\frac{V+21\text{mV}}{-8\text{mV}})} \\
\alpha_h &= \frac{0.0025}{1 + \exp(\frac{V+40\text{mV}}{8\text{mV}})} \\
\beta_h &= \frac{0.19}{1 + \exp(\frac{V+50\text{mV}}{-10\text{mV}})}.
\end{aligned} \tag{11}$$

The potassium channel (Doiron et al., 2001) model had maximum conductance $\bar{g}_K = 100\text{mS}$ and a reversal potential $E_K = -88.5$.

$$\begin{aligned}
I_m &= g_K m^3 h (V - E_K) \\
m_\infty &= \frac{1}{1 + \exp(\frac{V}{-19\text{mV}})} \\
\tau_m &= 0.8 \\
h_\infty &= \frac{1}{1 + \exp(\frac{V+3\text{mV}}{40\text{mV}})} \\
\tau_h &= 1.5.
\end{aligned} \tag{12}$$

Discussion

Here we introduced three important biological constraints to models of hierarchical dendritic computation. For each of the constraints, we asked how it impacted the model’s computational performance on several machine learning tasks. We found that the NaCaK nonlinearity performs as well if not better than the commonly used deep learning activation functions ReLU, LReLU, and sigmoid nonlinearities. Implementing synapse nonlinearities while using the NaCaK nonlinearity does not worsen binary tree model performance. When using this nonlinearity, constraining the weight parameters to be non-negative can greatly penalize performance in low multi-synaptic input repetition regimes, but implementing synapses along

with this constraint rescues model performance. The combination of all three constraints makes a model that performs as well as if not better than models without constraints. These constraints capture many limitations of real dendrites, suggesting that neurons with real dendrites may be able to actually solve the kinds of complex machine learning problems used in the deep learning field.

Limitations

Our dendritic binary tree model is technically a feed-forward ANN, which is an abstraction away from the temporally dynamic nature of neurons (Hines and Carnevale, 1997). In order to allow our model to work within this feed-forward context, we used a steady-state synaptic voltage nonlinearity approximation instead of real temporal conductance dynamics. Important neuronal phenomena, like backpropagating action potentials or dendritic direction selectivity (London and Häusser, 2005) cannot be reproduced using this model. In addition, inputs to neurons have a temporal dimension as well, so the input we use for this feed-forward model is not biologically plausible. An addition of real recurrent dynamics could render our model similar to those often implemented in compartmental modeling (Hines and Carnevale, 1997; Bower et al., 2013). While it would make conceptualization harder, implementing recurrent connections and using input with a time dimension could produce a more biologically plausible model and could be a very interesting constraint to add to the model.

The voltage-gated ion channel derived function, the NaCaK function, is a linear weighted sum of sodium, calcium, and potassium I-V curves where the weights of the sum are analogous to maximal conductance, which is a function of ion channel density. A limitation of this conception of dendritic nonlinearity is that we chose only 3 out of many ion channels that are present in a dendritic tree of any one neuron. We chose representative sodium (Hodgkin and Huxley, 1952), potassium (Doiron et al., 2001), and calcium (Miyasho et al., 2001) channels. It is an open question which ion channels would be best for making a general dendrite nonlinearity for the purposes of a model like the one we used. We also did not try to model any specific neuron type, so it would be very interesting to ask how distinct neuron types could support the computations we are analyzing. Follow-up work could be more specific about the kinds of neuron model they choose to constrain the model to.

Another limitation of this conception of dendritic nonlinearity is that the ratio of these ion channel density weights is, unrealistically, equal. Computational studies (Huys et al., 2006; Hay et al., 2011) have found that that not only is there no even ratio of these ion channels, this ratio is most certainly not consistent across the entirety of the morphology of the dendritic tree. Future work that allows these weights to be optimized might be fruitful for

possibly explaining ion channel density distributions in intricate dendritic morphologies. A combination of more realistic morphologies than the k -tree and a learnable NaCaK function for each node might introduce more biologically relevant degrees of freedom that could impact model computational performance.

High count multi-synaptic input repetitions to the terminals of separate dendritic sub-trees is a theoretical extreme of what real neurons are doing. These theoretical multi-synaptic boutons (MSB) assume that the repeated synaptic inputs go to different dendritic sub-trees. The literature shows support of same-branch MSBs and different-branch MSBs (Kincaid et al., 1998; Jones et al., 1997), but it is unclear if different-branch MSBs are to dendritic branches from completely different dendritic sub-trees. In addition, much evidence shows that dendritic trees are highly asymmetric (Ramon y Cajal, 1894), with synapses being at terminal nodes along lengths of dendritic node chains. It has been shown that these asymmetric trees are difficult to train on ML tasks using LReLU nonlinearities (Jones and Kording, 2021). Perhaps this work can inspire further work returning to this asymmetric tree architecture using the biological constraints we used in order to see if that allows the architecture to train better. In the context of the symmetric binary tree we use with multiple sub-trees, it may be more biologically relevant to look at performances of the low and medium repetition conditions.

The input-output functions from the deep learning literature that we use here to quantify computational capacity are a bad approximation to the problems solved by real neurons. The literal input-output (I/O) function of mapping an input of a 1-D, complex pixelated image to the presence or absence of an action potential is clearly biologically unrealistic. For example, in the visual system, the outputs of LGN neurons would have different structure (DiCarlo et al., 2012). However, we use this simple I/O function to explore the computational possibilities of an individual neuron. It would be interesting to see if the real functions computed by neurons are easier or harder.

In this study we used the backpropagation of error learning algorithm. Real neurons may learn this way or in entirely different ways and this is an active and controversial area of study (Lillicrap et al., 2020). A critique of the backpropagation of error algorithm is that it remains unobserved how the neuron achieves error feedback to the synapses, allowing the synapses to change based off of a gradient signal (Lillicrap et al., 2020). Alternatives to backpropagation of error, such as feedback alignment (Lillicrap et al., 2016), predictive coding (Millidge et al., 2020), or equilibrium propagation Scellier and Bengio (2017), could be used in this testing context. Further work may also show how plasticity rules in combination with third-factor signaling (Richards and Lillicrap, 2019) could possibly provide the error feedback a single neuron needs to adjust its synaptic weights. As for plasticity of dendritic branches, they

are most plastic during early neuron development (Koleske, 2013). Work may be needed to make models that distinguish between fully plastic neurons during early development and mature neurons which only have plastic synapses. Further work like this on neuron model learnability is needed in order to evaluate potential learning models in the future.

It is possible now to directly study dendritic integration in real neurons, although it is hard to do justice to their high-dimensional inputs (Spruston, 2009). One can record electrophysiologically dendrite-to-soma firing I/O functions that have the specific constraint of a multi-dendritic input corresponding to an action potential output. Further work using 2-photon microscopy, glutamate uncaging, dendritic voltage dyes, and electrophysiological recording could be useful in generating the kind of I/O function data that can produce further constraints on dendritic computation. It would be very interesting to calibrate neurons against physiological data and then ask about their computational limitations.

Contributions

We found that the NaCaK nonlinearity and swish nonlinearity performed the best in the sparse binary tree architecture. The NaCaK nonlinearity relates to the transfer function literature that has shown that smooth functions that are locally similar to ReLU tend to perform well (Xu et al., 2015). We also point out that the local properties of the neuron share some aspects of ResNets (Xie et al., 2017; He et al., 2016). Without local nonlinearities, a node simply copies its parents, producing the unity function. ResNets have the capacity to have better performances in machine learning benchmark tasks than a traditional multi-layer perceptron (Xie et al., 2017). Also, the nonlinear component of the NaCaK function contributes gradients (Figure S1) used by the backpropagation of error algorithm used to train the model, allowing the neuron model to learn the binary classification task more effectively than a completely linear model (See and compare performances to Table 1). These components of the NaCaK function show that producing learning systems that are more similar to those of real neurons may benefit from our considerations, e.g. for the design of specialized hardware. Interestingly, the sigmoid nonlinearity generally did worse than the NaCaK nonlinearity, suggesting that the brain has evolved a good transfer function.

The three constraints we implemented either did not worsen or improved the performance of our dendritic binary tree model in mid-to-high multi-synaptic repetition regimes. In low repetition regimes, binary trees using NaCaK nonlinearity and the non-negative weight constraint suffered in performance accuracy without the synapse nonlinearity. We suggest two potential reasons for why the synapse nonlinearity rescued performance. First, the synapse nonlinearity significantly expands the model size by introducing 4 more parameters per in-

put. The synapse nonlinearity merely represents a few degrees of freedom that exist in a real synapse. (Koch, 1999) This increase in synapse complexity in contrast to the more simple traditional way of representing synaptic strength with a scalar could be a reason why performance is rescued in this case. Second, it is important to note that the synapse nonlinearity maps the input, which ranges from 0 to 1, to synaptic voltage activation in millivolts, and the change in range of the NaCaK nonlinearity is across a wide domain between -70 and 50 mV (Figures S1, S6, and S7). Without the synapse nonlinearity, the activation in the binary tree would likely only sit close to 0. Having both the range of activations between -70 and 50 that the synapse nonlinearity enforces and the NaCaK nonlinearity with a matching range likely takes full advantage of the nonlinearities as activation travels down the binary tree. Though it is still unclear why non-negative weight constrained models need this matching in order to perform well, this may be why the combination of all three constraints leads to consistent or better performance.

Using the biological constraints we introduced a variety of free parameters analogous to properties in real neurons. Upon further analysis we found that in a model trained on MNIST with all three biological constraints the synaptic axial conductance weights (Figure S7) were close to the ranges found in Araya et al. (2014). The dendritic axial conductance weights (Figure S8A) were similar in magnitude. The resultant activations were also within the expected range of between -70 and 50 mV (Figures S7 and S8B,C). These results convince us that the model, when given biological constraints, yields biologically relevant behavior.

We take the time here to discuss the term "constraint." Given that our formulation of the synapse nonlinearity technically expands the size of our model, it may be seen as inappropriate to call the synapse nonlinearity a constraint. Interestingly, the machine learning field will tend to view the number of parameters as main constraint (Schölkopf et al., 2002) while the neuroscience field may rather view the existence of the channels that exist with their biological properties as a constraint - after all there are simply computations that are impossible under the biological constraints (e.g. those where the membrane voltage is $>100\text{mV}$). As such, we do not agree with just viewing the number of parameters as a constraint. Before constructing a synapse in a model and before knowing how that construction impacts model performance, we would be agnostic about whether the property limits or enhances model performance. In addition, we call this property a constraint because of its grounding in observed reality that abstracted models do not include or engage with. Therefore we call the synaptic nonlinearity a constraint in reference to its grounding or "constraining" our model to biological realism.

We introduced three biological constraints to an abstracted computational model of a neuron with dendrites, and observed the impacts of these constraints on computational per-

formance on well-defined ML tasks. We found that NaCaK nonlinearity which approximates real dendrite physiology clearly outperforms other nonlinearities. The addition of a synapse nonlinearity and non-negative weight constraint does not worsen performance on ML tasks, and in some cases improves performance. This more biologically plausible model can be trained on well-defined tasks, which opens further possibilities of exploring what neurons are capable of computationally. Our findings contribute to the theoretical evidence that dendritic computation can approximate complex nonlinear I/O functions, even with the constraints imposed by the components of dendrites.

Acknowledgments

We would like to acknowledge the members of the Kording Lab, specifically Roozbeh Farhoodi, Ben Baker and Ari Benjamin, for help in the development of this project. This work was funded by grants from the National Institute of Health, National Science Foundation, and Howard Hughes Medical Institute.

Code

The code for this project can be found at the following github repository:

https://github.com/ilennaj/ktree_constraints

Bibliography

- Agmon-Snir, H., Carr, C. E., and Rinzel, J. (1998). The role of dendrites in auditory coincidence detection. *Nature*, 393:268–272.
- Antic, S. D., Zhou, W. L., Moore, A. R., Short, S. M., and Ikonomu, K. D. (2010). The decade of the dendritic NMDA spike. *Journal of Neuroscience Research*, 88(14):2991–3001.
- Araya, R., Vogels, T. P., and Yuste, R. (2014). Activity-dependent dendritic spine neck changes are correlated with synaptic strength. *Proceedings of the National Academy of Sciences of the United States of America*, 111(28).
- Bower, J. M., Cornelis, H., and Beeman, D. (2013). *GENESIS, The GEneral NEural SIMulation System*, pages 1–8. Springer New York, New York, NY.
- Branco, T. and Häusser, M. (2010). The single dendritic branch as a fundamental functional unit in the nervous system.
- Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., and Ha, D. (2018). Deep Learning for Classical Japanese Literature. *Advances in Neural Information Processing Systems*, pages 1–8.
- Cohen, G., Afshar, S., Tapson, J., and Van Schaik, A. (2017). EMNIST: Extending MNIST to handwritten letters. *Proceedings of the International Joint Conference on Neural Networks*, 2017-May:2921–2926.
- David, B., Idan, S., and Michael, L. (2019). Single Cortical Neurons as Deep Artificial Neural Networks. *bioRxiv*.
- DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3):415–434.
- Doiron, B., Turner, R. A. Y. W., Maler, L., and Turner, R. W. (2001). Model of Gamma Frequency Burst Discharge Generated by Conditional Backpropagation. *Journal of Neurophysiology*, 86(4):1523–1545.
- Gidon, A., Zolnik, T. A., Fidzinski, P., Bolduan, F., Papoutsis, A., Poirazi, P., Holtkamp, M., Vida, I., and Larkum, M. E. (2020). Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science (New York, N.Y.)*, 367(6473):83–87.

- Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., and Shet, V. (2014). Multi-digit number recognition from street view imagery using deep convolutional neural networks. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, pages 1–13.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The elements of statistical learning*. Springer-Verlag.
- Hay, E., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2011). Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. *PLoS Computational Biology*, 7(7).
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:770–778.
- Hines, M. L. and Carnevale, N. T. (1997). The NEURON simulation environment. *Neural computation*, 9(6):1179–209.
- Hodgkin and Huxley (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiology*, 1117:500–544.
- Huys, Q. J. M., Ahrens, M. B., and Paninski, L. (2006). Efficient Estimation of Detailed Single-Neuron Models. *Journal of Neurophysiology*, 96(2):872–890.
- Izhikevich (2007). *Dynamical Systems in Neuroscience*. MIT Press.
- Jones, I. S. and Kording, K. P. (2021). Might a single neuron solve interesting machine learning problems through successive computations on its dendritic tree? *Neural Computation*, *In Press*.
- Jones, T. A., Klintsova, A. Y., Kilman, V. L., Sirevaag, A. M., and Greenough, W. T. (1997). Induction of multiple synapses by experience in the visual cortex of adult rats. *Neurobiology of Learning and Memory*, 68(1):13–20.
- Kincaid, A. E., Zheng, T., and Wilson, C. J. (1998). Connectivity and convergence of single corticostriatal axons. *Journal of Neuroscience*, 18(12):4722–4731.
- Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, New York.

- Koch, C., Poggio, T., and Torre, V. (1982). Retinal ganglion cells: a functional interpretation of dendritic morphology. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 298(1090):227–263.
- Koleske, A. J. (2013). Molecular mechanisms of dendrite stability. *Nature Reviews Neuroscience*, 14(8):536–550.
- Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. *ArXiv*.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 8:713–772.
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *proc. of the IEEE*.
- Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7:1–10.
- Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., and Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6):335–346.
- London, M. and Häusser, M. (2005). Dendritic Computation. *Annual Review of Neuroscience*, 28(1):503–532.
- Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 28.
- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.
- Mel, B. (2016). Toward a simplified model of an active dendritic tree. In Stuart, G. J., Spruston, N., and Häusser, M., editors, *Dendrites*. Oxford Scholarship Online.
- Millidge, B., Tschantz, A., and Buckley, C. L. (2020). Predictive coding approximates back-prop along arbitrary computation graphs. *arXiv*, pages 1–25.
- Minsky, M. and Papert, S. (1969). *Perceptrons*. MIT Press.

- Miyasho, T., Takagi, H., Suzuki, H., Watanabe, S., Inoue, M., Kudo, Y., and Miyakawa, H. (2001). Low-threshold potassium channels and a low-threshold calcium channel regulate Ca²⁺ spike firing in the dendrites of cerebellar Purkinje neurons: A modeling study. *Brain Research*, 891(1-2):106–115.
- Moldwin, T. and Segev, I. (2019). Perceptron learning and classification in a modeled cortical pyramidal cell. *bioRxiv*, page 464826.
- Nair, V. and Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on Machine Learning*.
- Poirazi, P., Brannon, T., and Mel, B. W. (2003a). Arithmetic of subthreshold synaptic summation in a model CA1 pyramidal cell. *Neuron*, 37(6):977–987.
- Poirazi, P., Brannon, T., and Mel, B. W. (2003b). Pyramidal neuron as two-layer neural network. *Neuron*, 37(6):989–999.
- Poirazi, P. and Papoutsis, A. (2020). Illuminating dendritic function with computational models. *Nature Reviews Neuroscience*, 21(6):303–321.
- Rall, W. (1964). Theoretical Significance of Dendritic Trees for Neuronal Input-Output Relations. In R F Reiss, editor, *Neural Theory and Modeling*, number February, chapter 4.2. Stanford University Press.
- Rall, W. (1967). Distinguishing Theoretical Synaptic Potentials Computed for Different Soma-Dendritic Distributions of Synaptic Input. *J Physiology*.
- Ramon y Cajal, S. (1894). The Croonian lecture. — The Fine Structure of Nervous Centers. *The Royal Society*, 55(331-335).
- Richards, B. and Lillicrap, T. P. (2019). Dendritic solutions to the credit assignment problem. *Current Opinion in Neurobiology*, 54:28–36.
- Scellier, B. and Bengio, Y. (2017). Equilibrium Propagation: Bridging the Gap Between Energy-Based Models and Backpropagation. *Frontiers in Computational Neuroscience*, 11(May).
- Schölkopf, B., Smola, B., Smola, A. J., Bach, F., Press, M. I. T., and Scholkopf, M. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive computation and machine learning. MIT Press.
- Spruston, N. (2009). Dendritic signal integration. *Encyclopedia of Neuroscience*, 1:445–452.

- Xiao, H., Rasul, K., and Vollgraf, R. (2017). Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv preprint*, pages 1–6.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated residual transformations for deep neural networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:5987–5995.
- Xu, B., Wang, N., Chen, T., and Li, M. (2015). Empirical Evaluation of Rectified Activations in Convolutional Network. *arXiv*.

Supplementary Figures

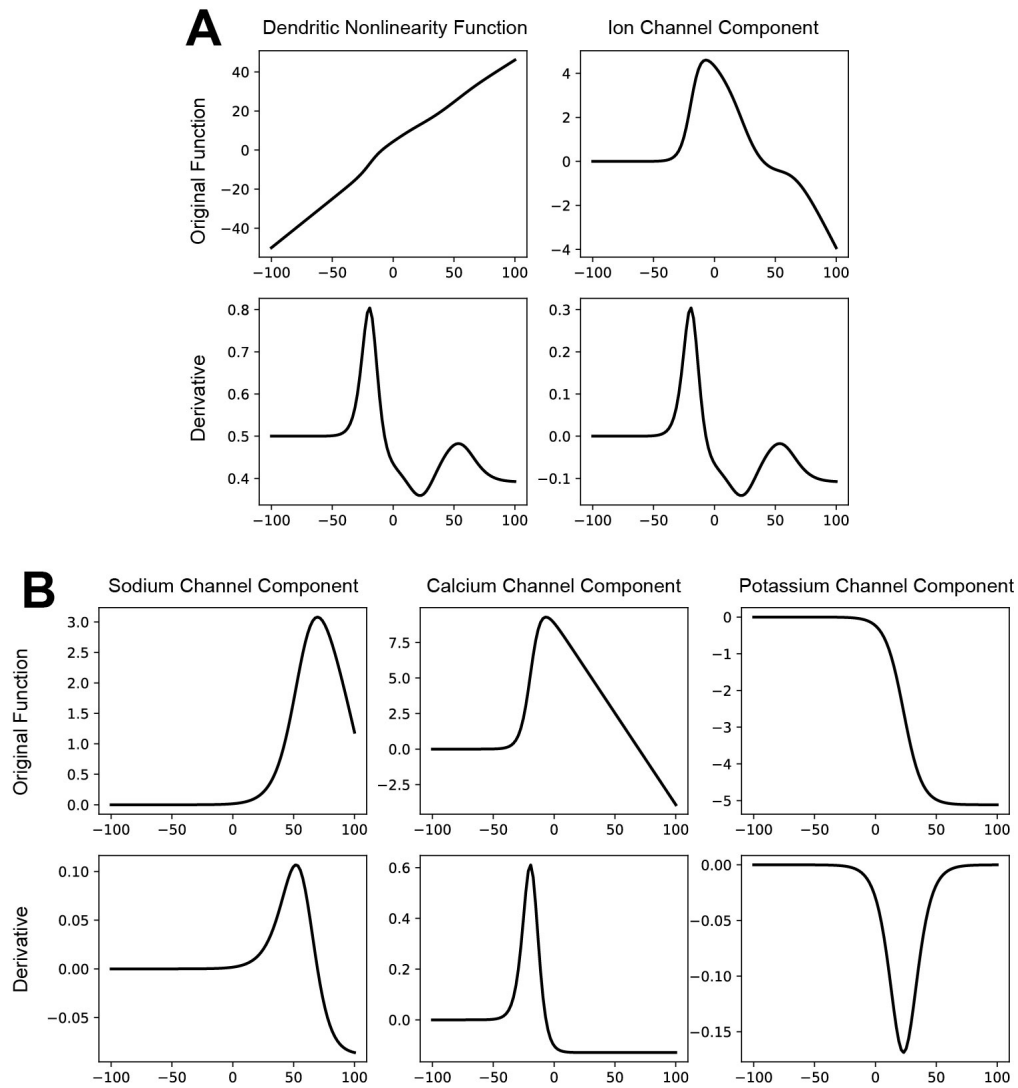


Figure S1: Dendritic Nonlinearity implemented as a linear combination of Sodium, Calcium, and Potassium Current-Voltage Curves. Dendritic Nonlinearity is called the NaCaK function. A) Top: Original NaCaK Function and its Ion channel component. Bottom: Derivatives of these functions. B) Top: Sodium, Calcium, and Potassium channel fitted I-V curves. Bottom: Derivatives of these I-V curves.

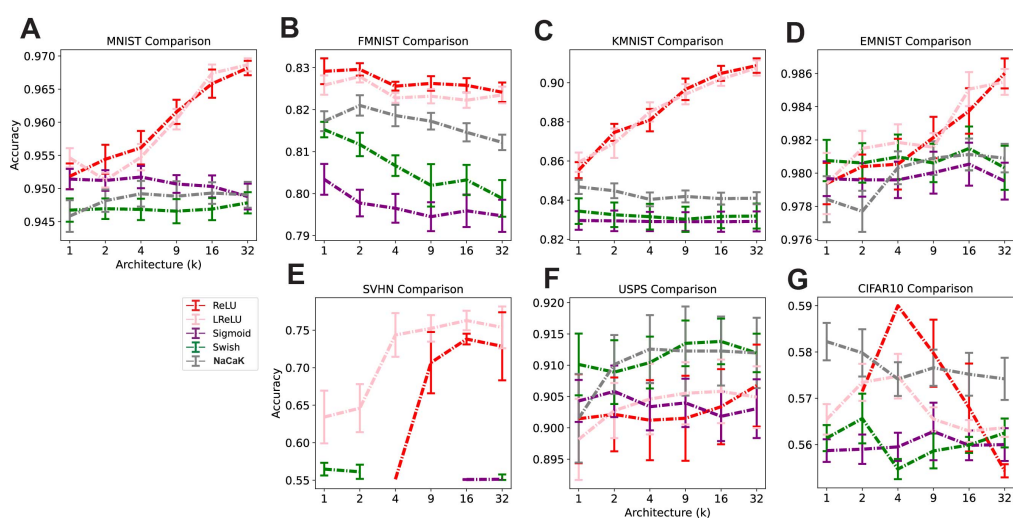


Figure S2: Comparing Nonlinear Activation Functions for the Dendrite Nodes in the FCNN For Each Dataset. Architecture ranges from low hidden layer width ($k=1$) to high hidden layer width ($k=32$) and mean accuracy of each model is displayed with standard error bars. Number of trials ranges from 1-10, and trial accuracy was omitted if the training was deemed to fail (accuracy threshold 0.55). Table 5 lists trial counts. Red: ReLU, Pink: LReLU, Purple: Sigmoid, Green: Swish, Grey: NaCaK Nonlinearity

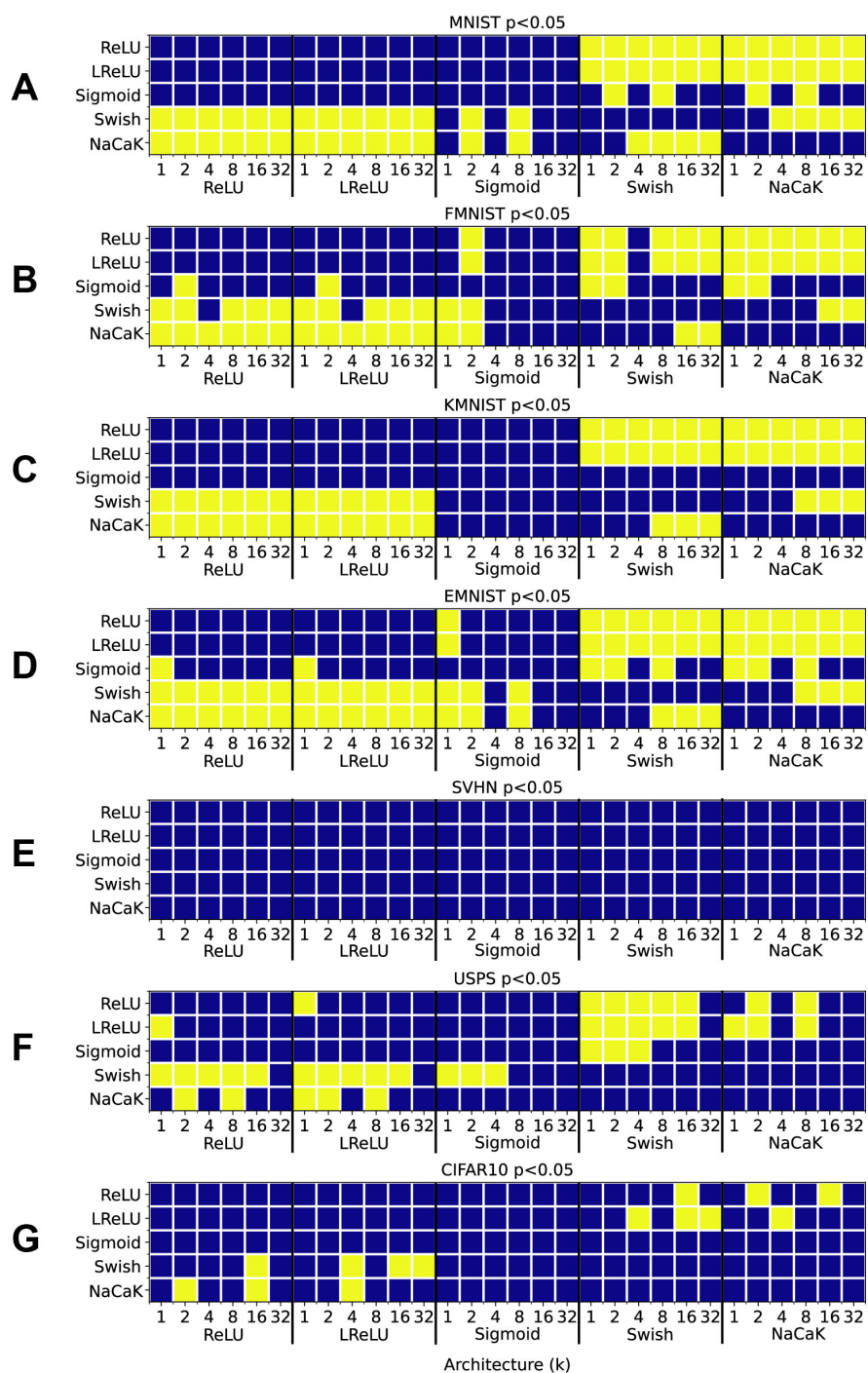


Figure S3: Significance Indicator Matrix for Comparing Nonlinear Activation Functions for the Dendrite Nodes in the k-tree. Yellow squares indicate $p < 0.05$ significance and blue squares indicate $p > 0.05$ significance. Architecture ranges from low multi-synaptic input repetitions ($k=1$) to high repetitions ($k=32$). Number of trials ranges from 1-10, and if there were not enough trials to run a t-test, then the block is automatically blue.

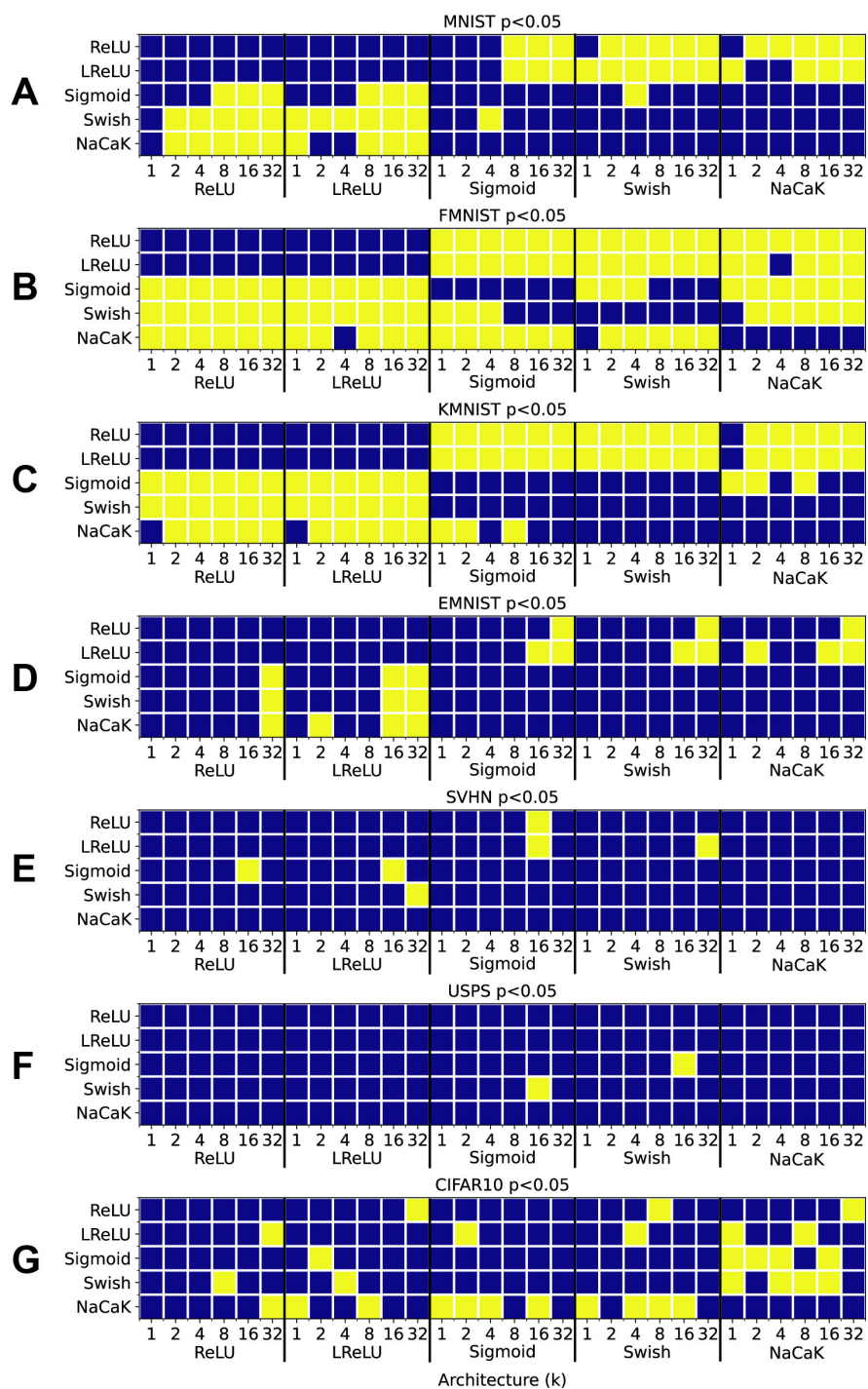


Figure S4: Significance Indicator Matrix for Comparing Nonlinear Activation Functions for the Dendrite Nodes in the FCNN. Yellow squares indicate $p < 0.05$ significance and blue squares indicate $p > 0.05$ significance. Architecture ranges from low hidden layer width ($k=1$) to high hidden layer width ($k=32$). Number of trials ranges from 1-10, and if there were not enough trials to run a t-test, then the block is automatically blue.

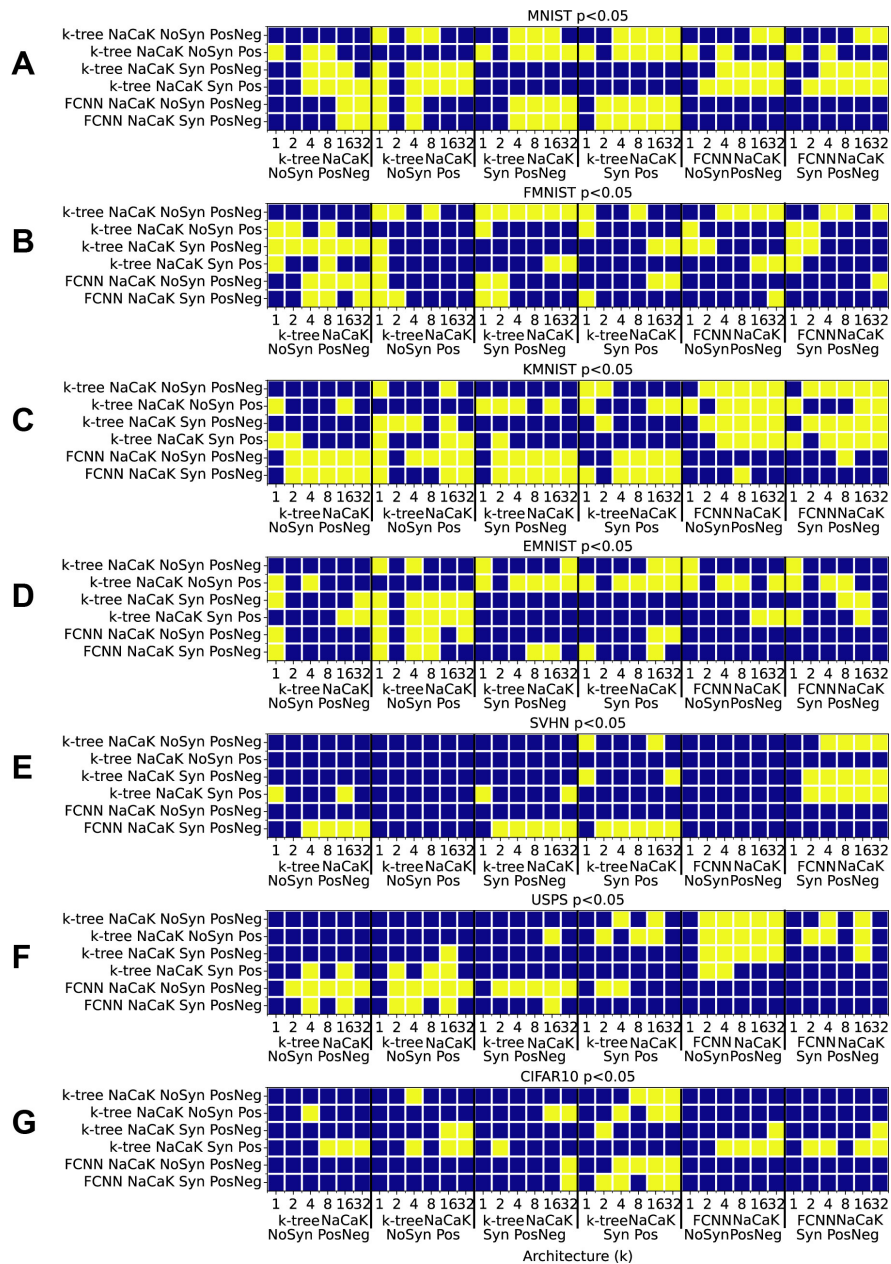


Figure S5: Significance Indicator Matrix for Comparing *k*-tree and FCNN Models With or Without Synapse Nonlinearity or Non-negativity Weight Constraint. Yellow squares indicate $p < 0.05$ significance and blue squares indicate $p > 0.05$ significance. Architecture ranges from low multi-synaptic input repetitions ($k=1$) to high repetitions ($k=32$). Number of trials ranges from 1-10, and if there were not enough trials to run a t-test, then the block is automatically blue. NoSyn = No Synapse Nonlinearity. Syn = Synapse Nonlinearity. PosNeg = Positive and Negative Weights. Pos = Non-negative Weights.

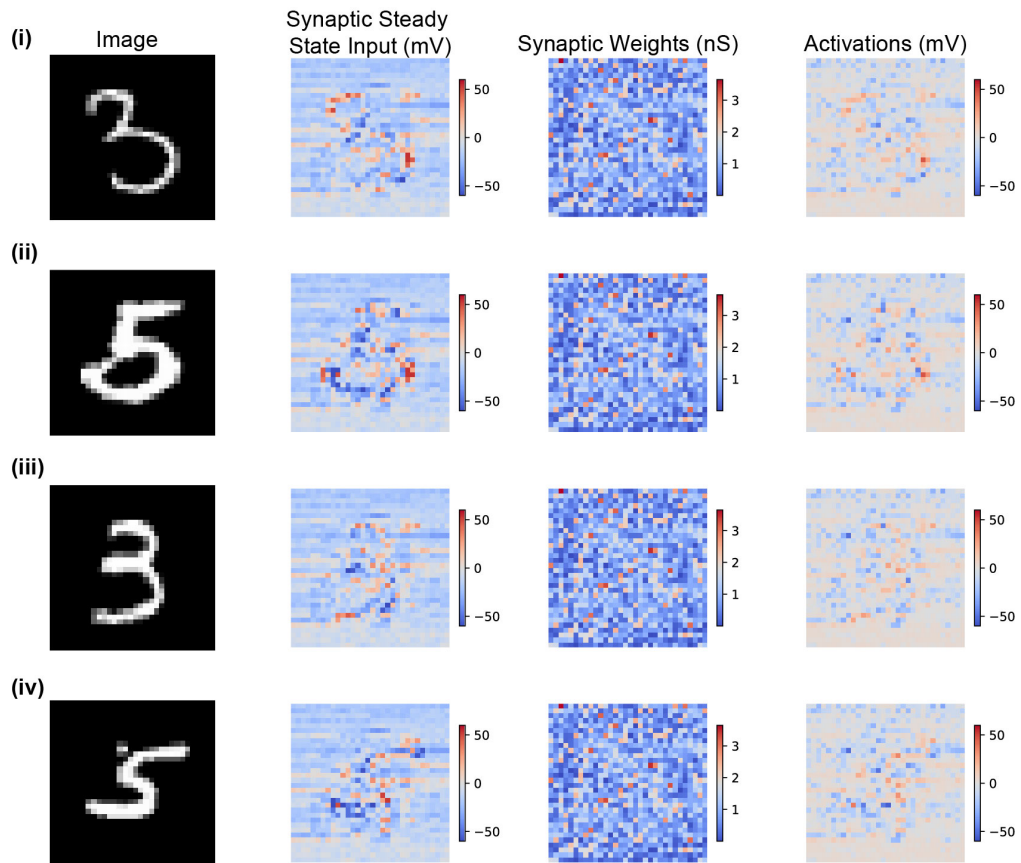


Figure S6: Heatmaps of Synaptic Activations. The images (left) are vectorized and input into a synaptic nonlinearity layer, which yields a synaptic steady state input between -70 and 50 mV. The product of the synaptic weights, analogous to synaptic bouton neck conductances, and the synaptic steady state inputs then lead a final synaptic activation (right) that is received by downstream dendrite nodes.

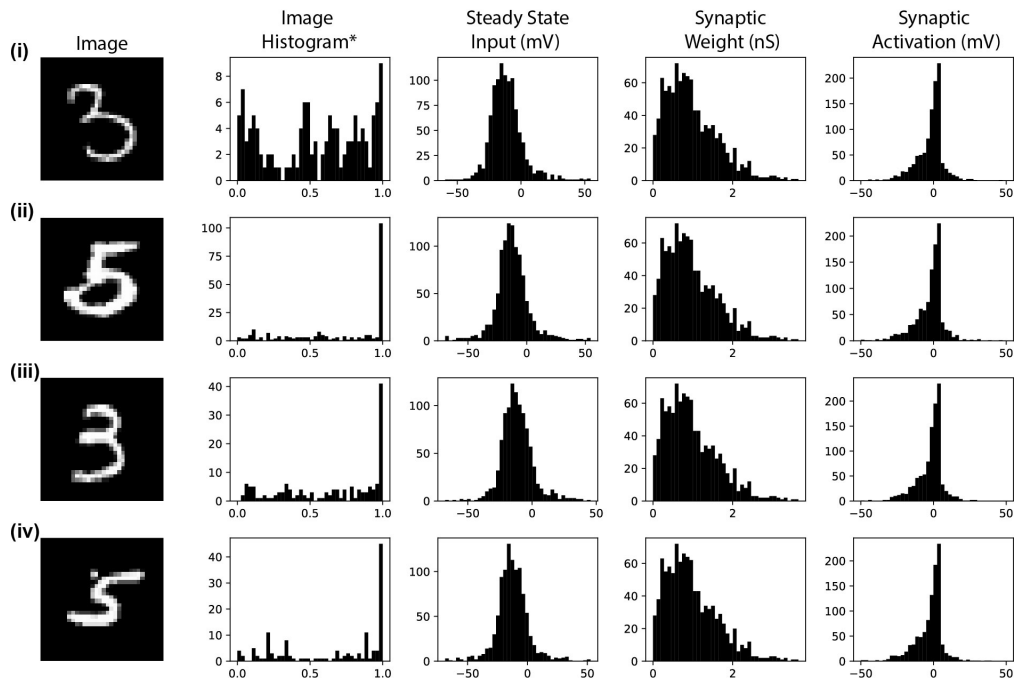


Figure S7: Histograms of Synaptic Activations. The images (left) are vectorized and input into a synaptic nonlinearity layer, which yields a synaptic steady state input between -70 and 50 mV. The product of the synaptic weights, analogous to synaptic bouton neck conductances, and the synaptic steady state inputs then lead a final synaptic activation (right) that is received by downstream dendrite nodes.

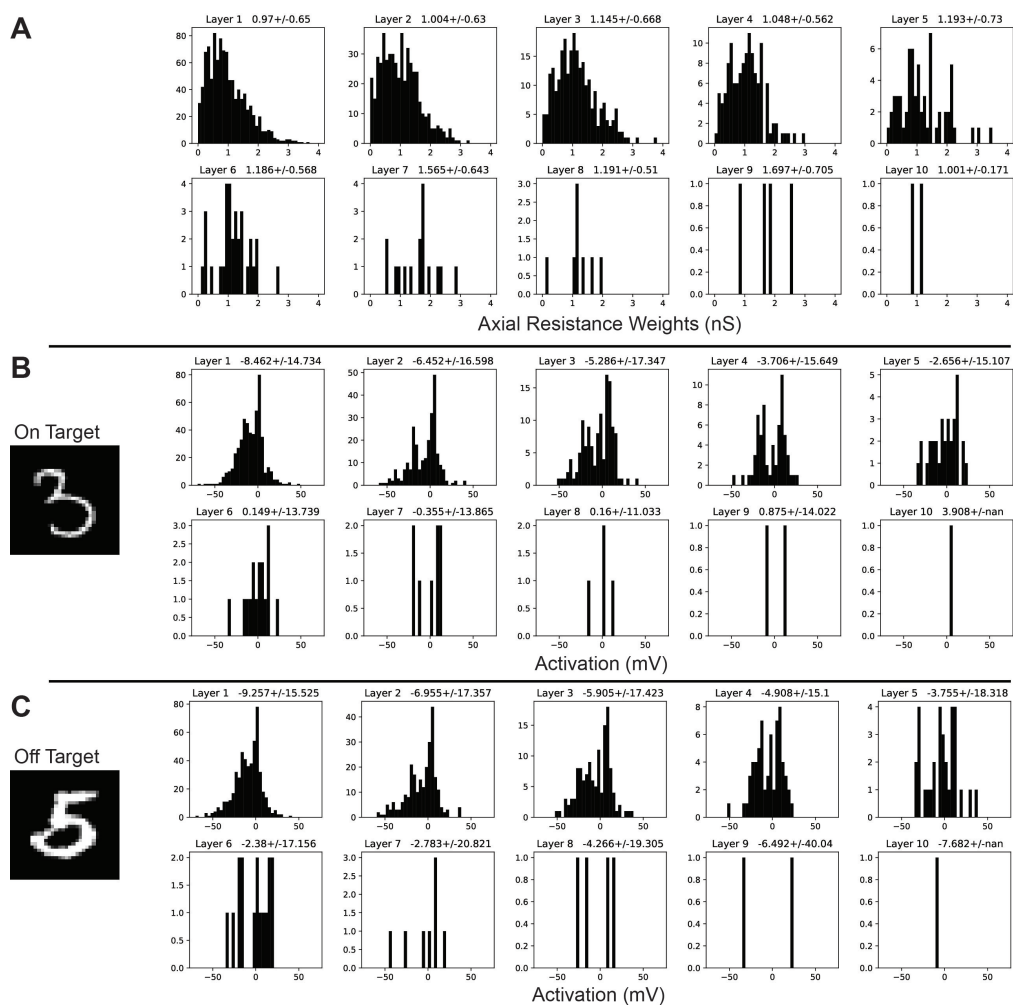


Figure S8: Distributions of Dendritic Node Weights and Activations of a Trained 1-tree. A) Dendritic node weight distribution (in nanoSiemens) in layers 1-10 of a 1-tree. Layers are labeled with layer number, mean and standard deviation. B and C) Dendritic node activation distribution (in milliVolts) in layers 1-10 of a 1-tree for "On Target" (B) and "Off Target" (C) stimuli. Layers are labeled with layer number, mean, and standard deviation.

MNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	8	9	10	9	10	10
LReLU	8	10	10	10	10	10
Sigmoid	1	5	0	2	0	0
Swish	10	10	10	10	10	10
NaCaK	10	10	10	10	10	10

FMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	7	10	8	10	10	10
LReLU	7	10	10	10	10	10
Sigmoid	2	3	0	0	0	1
Swish	4	7	8	9	10	10
NaCaK	9	10	10	10	10	10

KMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	8	9	9	10	10	10
LReLU	7	7	10	10	10	10
Sigmoid	1	1	1	0	0	0
Swish	10	10	10	10	10	10
NaCaK	9	10	10	10	10	10

EMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	10	10	10	10	10	10
LReLU	10	10	10	10	10	10
Sigmoid	8	3	1	2	0	0
Swish	10	10	10	10	10	10
NaCaK	10	10	10	10	10	10

SVHN						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	0	0	0	1	0	1
LReLU	1	1	1	1	1	4
Sigmoid	0	0	0	0	0	0
Swish	4	2	2	3	3	3
NaCaK	2	3	4	4	3	2

USPS						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	6	8	10	10	10	10
LReLU	6	9	10	10	10	10
Sigmoid	3	3	2	0	0	0
Swish	10	10	10	10	10	10
NaCaK	8	6	7	10	10	10

CIFAR10						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	2	4	4	5	5	6
LReLU	2	1	5	3	7	6
Sigmoid	0	0	0	0	0	0
Swish	1	4	3	2	4	6
NaCaK	10	9	8	8	9	8

Table 2: Trial Numbers for Figure 2, Comparing Nonlinear Activation Functions for the Dendrite Nodes in the k -tree For Each Dataset.

MNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree SQGL NoSyn PosNeg	10	10	10	10	10	10
k-tree SQGL Syn PosNeg	9	10	10	10	10	10
FCNN SQGL NoSyn PosNeg	10	10	10	10	10	10
FCNN SQGL Syn PosNeg	10	10	10	10	10	10
FMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree SQGL NoSyn PosNeg	9	10	10	10	10	10
k-tree SQGL Syn PosNeg	10	10	10	10	10	10
FCNN SQGL NoSyn PosNeg	10	10	10	10	10	10
FCNN SQGL Syn PosNeg	10	10	10	10	10	10
KMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree SQGL NoSyn PosNeg	9	10	10	10	10	10
k-tree SQGL Syn PosNeg	10	10	10	10	10	10
FCNN SQGL NoSyn PosNeg	10	10	10	10	10	10
FCNN SQGL Syn PosNeg	10	10	10	10	10	10
EMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree SQGL NoSyn PosNeg	10	10	10	10	10	10
k-tree SQGL Syn PosNeg	10	10	10	10	10	10
FCNN SQGL NoSyn PosNeg	10	10	10	10	10	10
FCNN SQGL Syn PosNeg	10	10	10	10	10	10
SVHN						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree SQGL NoSyn PosNeg	2	3	4	4	3	2
k-tree SQGL Syn PosNeg	9	10	10	10	10	10
FCNN SQGL NoSyn PosNeg	0	1	0	0	1	0
FCNN SQGL Syn PosNeg	7	10	10	10	10	10
USPS						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree SQGL NoSyn PosNeg	8	6	7	10	10	10
k-tree SQGL Syn PosNeg	10	10	10	10	10	10
FCNN SQGL NoSyn PosNeg	10	10	10	10	10	10
FCNN SQGL Syn PosNeg	10	10	10	10	10	10
CIFAR10						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree SQGL NoSyn PosNeg	10	9	8	8	9	8
k-tree SQGL Syn PosNeg	6	9	10	9	9	10
FCNN SQGL NoSyn PosNeg	8	8	10	10	10	10
FCNN SQGL Syn PosNeg	7	7	3	5	7	3

Table 3: Trial Numbers for Figure 3, Comparing k -tree and FCNN models with or without Synapse Nonlinearity.

MNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree NaCaK NoSyn PosNeg	10	10	10	10	10	10
k-tree NaCaK NoSyn Pos	5	9	10	10	10	10
k-tree NaCaK Syn PosNeg	9	10	10	10	10	10
k-tree NaCaK Syn Pos	10	10	10	10	10	10
FCNN NaCaK NoSyn PosNeg	10	10	10	10	10	10
FCNN NaCaK Syn PosNeg	10	10	10	10	10	10
FMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree NaCaK NoSyn PosNeg	9	10	10	10	10	10
k-tree NaCaK NoSyn Pos	4	10	10	10	10	10
k-tree NaCaK Syn PosNeg	10	10	10	10	10	10
k-tree NaCaK Syn Pos	10	10	10	10	10	10
FCNN NaCaK NoSyn PosNeg	10	10	10	10	10	10
FCNN NaCaK Syn PosNeg	10	10	10	10	10	10
KMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree NaCaK NoSyn PosNeg	9	10	10	10	10	10
k-tree NaCaK NoSyn Pos	6	8	10	10	10	10
k-tree NaCaK Syn PosNeg	10	10	10	10	10	10
k-tree NaCaK Syn Pos	10	10	10	10	10	10
FCNN NaCaK NoSyn PosNeg	10	10	10	10	10	10
FCNN NaCaK Syn PosNeg	10	10	10	10	10	10
EMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree NaCaK NoSyn PosNeg	10	10	10	10	10	10
k-tree NaCaK NoSyn Pos	10	10	10	10	10	10
k-tree NaCaK Syn PosNeg	10	10	10	10	10	10
k-tree NaCaK Syn Pos	10	10	10	10	10	10
FCNN NaCaK NoSyn PosNeg	10	10	10	10	10	10
FCNN NaCaK Syn PosNeg	10	10	10	10	10	10
SVHN						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree NaCaK NoSyn PosNeg	2	3	4	4	3	2
k-tree NaCaK NoSyn Pos	0	1	0	0	0	0
k-tree NaCaK Syn PosNeg	9	10	10	10	10	10
k-tree NaCaK Syn Pos	10	9	10	10	10	10
FCNN NaCaK NoSyn PosNeg	0	1	0	0	1	0
FCNN NaCaK Syn PosNeg	7	10	10	10	10	10
USPS						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree NaCaK NoSyn PosNeg	8	6	7	10	10	10
k-tree NaCaK NoSyn Pos	1	10	10	10	10	10
k-tree NaCaK Syn PosNeg	10	10	10	10	10	10
k-tree NaCaK Syn Pos	10	10	10	10	10	10
FCNN NaCaK NoSyn PosNeg	10	10	10	10	10	10
FCNN NaCaK Syn PosNeg	10	10	10	10	10	10
CIFAR10						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
k-tree NaCaK NoSyn PosNeg	10	9	8	8	9	8
k-tree NaCaK NoSyn Pos	0	1	4	2	3	3
k-tree NaCaK Syn PosNeg	6	9	10	9	9	10
k-tree NaCaK Syn Pos	9	10	10	10	9	10
FCNN NaCaK NoSyn PosNeg	8	8	10	10	10	10
FCNN NaCaK Syn PosNeg	7	7	3	5	7	3

Table 4: Trial Numbers for Figure 4, Comparing k -tree and FCNN models with or without Non-negative Weight Constraint.

MNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	10	10	10	10	10	10
LReLU	10	10	10	10	10	10
Sigmoid	10	10	10	10	10	10
Swish	10	10	10	10	10	10
NaCaK	10	10	10	10	10	10
FMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	7	10	10	10	10	10
LReLU	10	10	10	10	10	10
Sigmoid	10	10	10	10	10	10
Swish	10	10	10	10	10	10
NaCaK	10	10	10	10	10	10
KMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	10	10	10	10	10	10
LReLU	10	10	10	10	10	10
Sigmoid	10	10	10	10	10	10
Swish	10	10	10	10	10	10
NaCaK	10	10	10	10	10	10
EMNIST						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	10	10	10	10	10	10
LReLU	10	10	10	10	10	10
Sigmoid	10	10	10	10	10	10
Swish	10	10	10	10	10	10
NaCaK	10	10	10	10	10	10
SVHN						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	1	0	1	4	2	5
LReLU	4	6	8	6	7	9
Sigmoid	1	0	1	0	2	1
Swish	2	2	0	1	0	2
NaCaK	0	1	0	0	1	0
USPS						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	8	10	10	10	10	10
LReLU	10	10	10	10	10	10
Sigmoid	10	10	10	10	10	10
Swish	10	10	10	10	10	10
NaCaK	10	10	10	10	10	10
CIFAR10						
	1-tree	2-tree	4-tree	8-tree	16-tree	32-tree
ReLU	0	1	1	2	2	3
LReLU	9	9	9	10	7	8
Sigmoid	5	6	4	6	6	3
Swish	5	4	5	4	6	3
NaCaK	8	8	10	10	10	10

Table 5: Trial Numbers for Figure S2, Comparing Nonlinear Activation Functions for the Dendrite Nodes in the FCNN For Each Dataset.