# Dendritic Computing: Branching Deeper into Machine Learning

**Jyotibdha Acharya,** [a]* **Arindam Basu,** [b] **Robert Legenstein,** [c] **Thomas Limbacher,** [c] **Panayiota Poirazi** [d] **and Xundong Wu** [e12]

[a] *Institute of Infocomm Research, A\*STAR, Singapore*

[b] *Department of Electrical Engineering, City University of Hong Kong, Hong Kong*

[c] *Institute of Theoretical Computer Science, Graz University of Technology, Austria*

[d] *Institute of Molecular Biology and Biotechnology (IMBB), Foundation for Research and Technology-Hellas (FORTH), Greece*

[e] *School of Computer Science, Hangzhou Dianzi University, China*

Abstract—In this paper, we discuss the nonlinear computational power provided by dendrites in biological and artificial neurons. We start by briefly presenting biological evidence about the type of dendritic nonlinearities, respective plasticity rules and their effect on biological learning as assessed by computational models. Four major computational implications are identified as improved expressivity, more efficient use of resources, utilizing internal learning signals, and enabling continual learning. We then discuss examples of how dendritic computations have been used to solve real-world classification problems with performance reported on well known data sets used in machine learning. The works are categorized according to the three primary methods of plasticity used—structural plasticity, weight plasticity, or plasticity of synaptic delays. Finally, we show the recent trend of confluence between concepts of deep learning and dendritic computations and highlight some future research directions.

This article is part of a Special Issue entitled: SI: Dendrites. © 2021 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: non-linear dendrites, plasticity, rewiring, expressivity, maxout networks, machine learning, deep neural networks.

## 1. INTRODUCTION

The last decade has seen a resurgence of interest in neural networks fueled by the huge success of deep learning across different domains spanning object detection and recognition in images (Krizhevsky et al., 2012; He et al., 2016), natural language processing (Collobert et al., 2011), speech recognition (Hinton et al., 2012), face verification (Taigman et al., 2014) etc. Despite those achievements, there is no doubt that current deep neural network models are still far away from reaching the performance level of human intelligence. It is clear that biological neuronal networks are far superior to the state-of-the-art deep neural network models especially in sensorimotor tasks—a child can easily tie shoelaces while a robot would struggle with such fine manipulation. Quite naturally one therefore investigates biological neural networks in order to determine the missing parts.

The basic model of a neuron used in deep learning is still largely the same as the ones introduced by McCulloch and Pitts—linear synaptic summation followed by a lumped nonlinearity (the type of nonlinearity has been changed to rectified linear unit from sigmoid or Heaviside). This model totally ignores any possible computational role played by dendrites, the thin processes that extend from the cell bodies of neurons and serve as their primary receiving end. Processing of synaptic inputs by dendrites depends on various factors: the morphology of the dendritic tree, the composition of ionic mechanisms, and the plasticity rules that govern learning. These factors influence dendritic integration in interacting ways, and furnish dendrites with the ability to perform complex computations. Here, the term computation is used in a generic sense to refer to how well a dendrite can implement a given input–output mapping such as a logical AND/OR/XOR operation or implement a non-linear transformation such as signal amplification. These types of local computations help neurons to solve more complex problems like the discrimination of incoming patterns/memories.

Morphologically complex dendrites can, for example, act as filters due to the passive attenuation and temporal filtering of input signals traveling to the cell body (Mengual et al., 2020; Spruston, 2008), thus enabling neurons to infer the location and kinetics of incoming signals. Dendrites are also home to a wide variety of ionic mechanisms, many of which have voltage-gated (active) conductances that support the generation
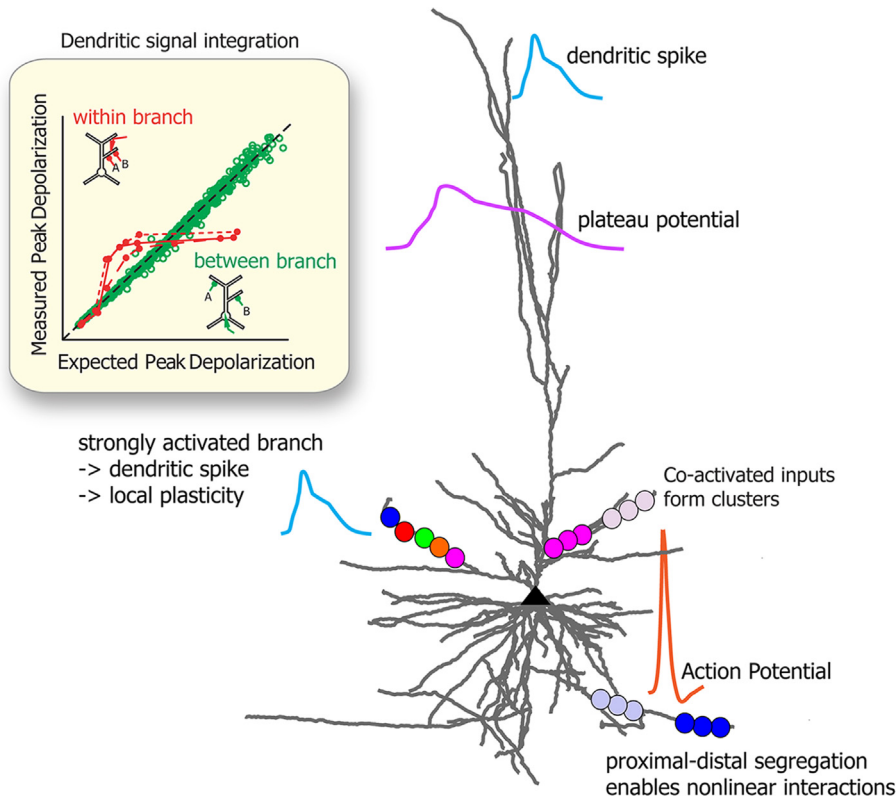
---

**Fig. 1.** Biological neurons support a variety of dendritic spikes, which can occur throughout the dendritic tree. They also support longer-lasting dendritic plateau potentials, which are primarily restricted to the apical tuft of pyramidal neurons. These regenerative events influence signal integration in non-linear ways. For example, integration of inputs within a strongly activated branch can be described by a sigmoidal activation function whereas the same inputs distributed across branches summate linearly (figure adopted from Polsky et al., 2004). Moreover, active dendrites implement a range of plasticity rules. Inputs that fire together tend to wire together, forming synapse clusters within individual branches. Inputs can also organize in distal vs. proximal ones within individual branches, thus enabling non-linear interactions between them. Note that different colors of synapses are used to indicate different input afferents that fire out of sync. Same color denotes afferents that fire synchronously.

of dendritic spikes (reviewed in (Major et al., 2013)), see Fig. 1. These spikes allow neurons to perform various nonlinear computations, ranging from logical operations to low-pass filtering, amplification or segregation of signals, coincidence detection, etc. (Ariav et al., 2003; Softky, 1994; Spruston, 2008; Payeur et al., 2019). The benefits provided by nonlinear dendritic integration are maximized through their coupling to a set of multiscale plasticity rules (Bono et al., 2017; KKastellakis and Poiraziastellakis and Poirazi, 2019) involving weight, delay, or structural plasticity (used interchangeably with rewiring in this paper) that underlie learning in biological networks.

The goal of this review paper is to summarize efforts to link theories of information processing in neural networks with nonlinear dendritic components with modern deep learning approaches. We first review prior work that has explored principles of biological learning in simulated neuronal networks with non-linear dendrites (Section 2) and highlight four main computational implications of having dendrites. However, the work summarized in this section does not demonstrate the practical feasibility of the above approaches in solving

real world problems such as image recognition where machine learning methods are typically used. Section 3 bridges this gap by reviewing studies that demonstrate how neurons with non-linear dendrites perform on traditional machine learning benchmark tasks and compare the performance with conventional algorithms. Finally, connections between deep learning and dendritic properties are summarized in Section 3.2, given the increased interest in this topic.

## 2. PRINCIPLES OF BIOLOGICAL LEARNING WITH NON-LINEAR DENDRITES

The role of nonlinear dendritic mechanisms in advancing neuronal computations has been investigated in various theoretical studies. The proposed computational implications can be broadly categorized into four groups:

- (P1) Improved expressivity of single neurons
- (P2) Improved use of neuronal resources and generalization capabilities
- (P3) Utilization of internal learning signals
- (P4) Enabling continual learning.

In the following subsections, we discuss example studies for each of these proposed computational advantages of neurons with nonlinear dendrites as opposed to point neuron models.

### 2.1. P1: Improved expressivity of single neurons

Due to the nonlinear dendritic properties of many neuron types, and in particular pyramidal cells, their integrative properties cannot be accurately described by the linear point neuron model. In fact, it was shown in (Poirazi et al., 2003) that the firing rate of a detailed compartmental model of a hippocampal CA1 pyramidal cell can be well-modeled by a two-layer neural network, see Fig. 2A. In this network the sigmoidal hidden layer represents nonlinear properties of the dendrites and the output neuron models the summation of dendritic currents at the soma. The two-stage model was also established viable for prediction of spike responses (Naud et al., 2014), for fast spiking interneuron models (Tzilivaki et al., 2019), and even for single human neurons (Gidon et al., 2020) (see (Jadi et al., 2014) for a historical review, see also (Beniaguev et al., 2020)).
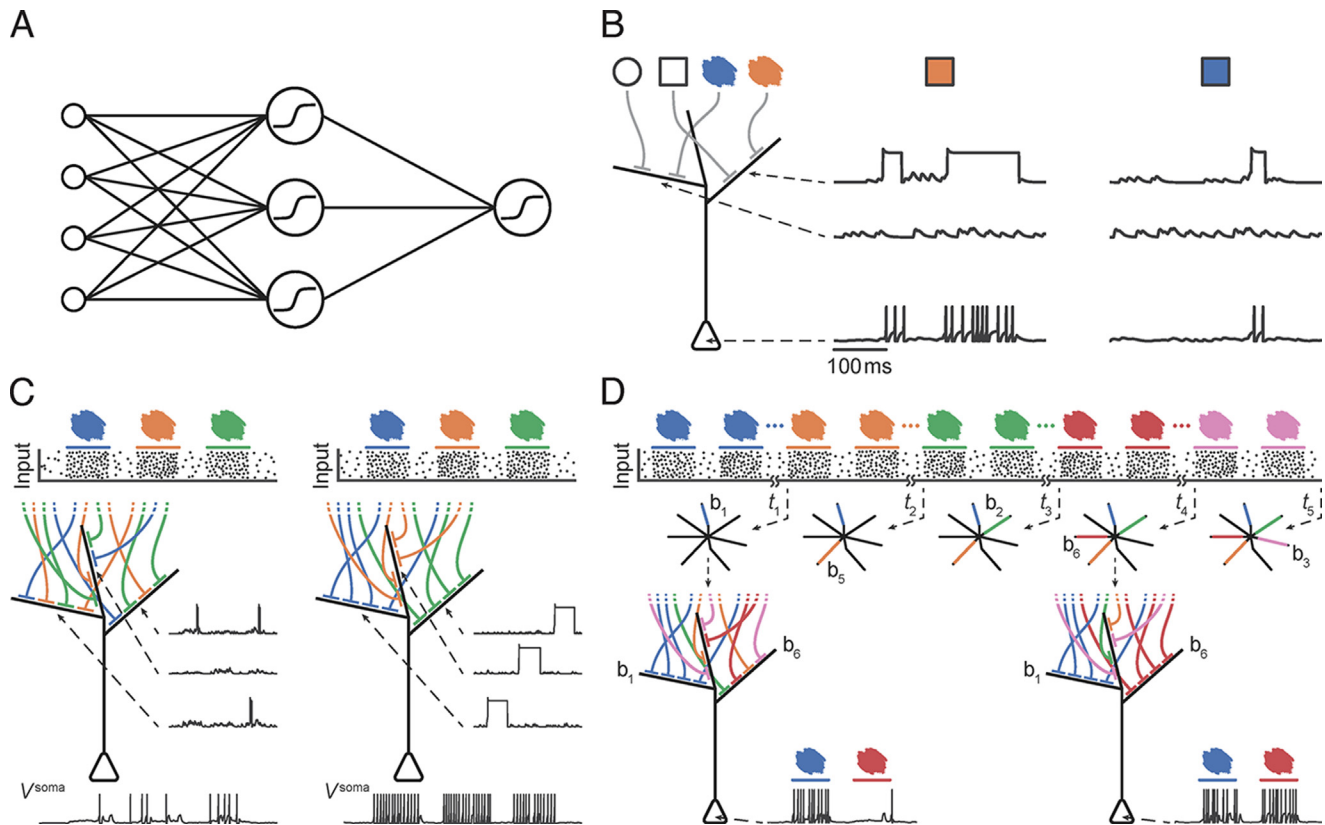
**Fig. 2.** Models of neurons with non-linear dendrites and possible computational implications. (**A**) A generic representation of a two-layer neuron model. Synaptic inputs (left) are integrated in non-linear and electrotonically segregated dendritic subunits. In the soma (right), dendritic outputs are added. In rate-based models, the neuron is equivalent to a two-layer artificial neural network. Typically, the inputs are sparsely connected to the dendritic subunits. Here, we depict the more general case of dense connections between these layers, which is in principle possible since axons can contact several branches of a neuron. (**B**) Two-layer model with dynamic dendritic plateau potentials. Segregated dendrites integrate synaptic inputs temporally. Temporally extended (on the order of 100 ms) dendritic plateau potentials are stochastically elicited at times of high branch potentials. Illustration of the feature binding problem. Abstract input features ('disk', 'square', 'blue', 'orange') are indicated by synaptic inputs. Since dendritic plateau potentials that lead to somatic action potentials are elicited based on the local potential, they appear only if feature combinations are active on a common branch (here: 'blue disk' and 'orange square') but not at different branches (here, 'blue square' and 'orange disk'). (**C**) Synaptic rewiring on dendritic branches. Abstract input features ('blue', 'orange', 'green') are encoded by a specific firing pattern of the input neurons (top). Initially, input neurons are connected to branches randomly, brief dendritic spikes appear occasionally (middle left), and the neuronal firing rate is low (bottom left). After rewiring, functionally related synapses are clustered on dendritic branches. Hence, the activation of the input features leads to long lasting dendritic plateau potentials (middle right) and to an elevated firing rate of the neuron (bottom right). (**D**) Rewiring protects stored information by segregating functionally unrelated inputs onto different dendritic branches. Here, abstract input features ('blue', 'orange', 'green', 'red', 'rose') are activated sequentially (top). Novel input features cluster on free branches while previously established clusters are retained (middle; shown is a schematic drawing of seven branches $b_i$ at five different time points). The neuron successively reorganizes its synaptic connections such that it responds to any input feature that has been presented previously (bottom; shown is the wiring diagram for three branches and the somatic response to two input features at two different time points).

For one particular setting of its synaptic weights, a neuron implements one function, i.e., one mapping from inputs to its outputs. For different weight settings, different functions can be implemented. The term "expressivity" of a neuron — or more generally of a parameterized model in machine learning — refers to the number of functions that can be implemented with different weight (parameter) settings. The larger the number of implementable functions, the more expressive is the model. If single neurons operate like two-layer networks as proposed by the above studies, it means that they are more expressive than point neuron models. The computational implications of the two-stage model have been investigated in several studies and with models of various granularity, from simple rate-

based models (Cazé et al., 2013) to more elaborated models with temporal dynamics of dendritic spikes (Limbacher and Legenstein, 2020), see Fig. 2. These studies show that single neurons can in principle compute linearly non-separable functions, which is impossible for point neuron models. A recent example is the ability of human neurons to solve the Exclusive-OR (XOR) problem at the dendritic level. Experiments in human L2/3 pyramidal neurons revealed a new type of dendritic spikes (termed dCaAPs), with a counter intuitive feature: their amplitude decreases with stimulus strength. Biophysical modeling showed that the presence of these dCaAPs enables individual neurons—in fact individual dendrites—to solve the XOR problem, a computation believed solvable only by multi-layer neural networks

(Gidon et al., 2020). Another example for a linearly non-separable function is the feature binding problem (Legenstein and Maass, 2011) shown in Fig. 2B. Consider four synaptic inputs to a neuron indicating abstract features of a visual input. The features are the geometric shape ('disk' or 'square') and the color ('blue' or 'orange'). The neuron should be active when a blue disk or an orange square is encountered, but not for a orange disk or a blue square. This is impossible for a point neuron model, but easy to implement with a two-stage model (Fig. 2B). In addition, it is also shown in (Legenstein and Maass, 2011) that local plasticity rules can self-organize the synaptic connections to learn such feature bindings just from exposure to the corresponding patterns.

The role of local plasticity rules in maximally engaging dendritic nonlinearities was also nicely exemplified by Poirazi and Mel (Poirazi and Mel, 2001) on a binary classification task using simplified mathematical neurons. In that work, the authors trained simplified (star-like shaped) neurons (with $m$ branches (subunits) and $k$ synapses per branch) under two schemes of dendritic integration: (1) a point-neuron like setting, where all weighted synaptic inputs were combined linearly before passing the somatic nonlinearity and (2) a nonlinear setting, where each branch had it's own nonlinear activation function and branch outputs were summed before passing the somatic nonlinearity.

Learning involved both weight updates as well as structural (rewiring) changes, as outlined in Table 1. Plasticity was driven by the degree of synchronous activation between synapses of the same branch. The authors showed that this novel structural learning rule, allowed a cell with nonlinear branches to learn $> 45$ times more input patterns than a cell with linear branches, despite both cell types having exactly the same number of trainable weights. The increase in storage capacity was also demonstrated analytically, by counting all possible input configurations (originating from a given number of axons) that led to district outcomes for each cell type. This finding also indicates a more efficient use of synaptic resources by nonlinear dendrites, a topic that is described in more detail in Section 2.2.

Of note, while this study did use local information (the activity correlation among synapses within a branch), it also relied on a global instructive signal (the cell's response error) to guide plasticity within dendrites. The latter is similar to how artificial neural networks (ANNs) learn via minimizing a global loss function. Here, the fitness score $C$ needed the calculation of derivatives of dendritic and somatic activations, which are computationally complex. Section 3.1 shows several examples of how modified versions of this learning rule with reduced complexity can be used for solving benchmark problems conventionally used in the machine learning community. These simplifications also allow power-efficient integrated circuit realizations of the same using neuromorphic approaches (Bhaduri et al., 2018).

Lastly, nonlinear dendritic integration is not only beneficial for pattern recognition or memory. In a computational study, it was proposed that nonlinear dendritic computations could enable neuronal networks to perform sampling-based probabilistic inference on a quite general class of probability distributions (Pecevski et al., 2011).

### 2.2. P2: Improved use of neuronal resources and improved generalization capabilities

While single neurons are clearly more expressive with nonlinear dendrites, networks of point neuron models are universally expressive even without dendritic nonlinearities. Specifically, it is well-known that even two-layer artificial neural networks are universal approximators. Namely, for any given function on a compact domain and a given demanded precision, one can construct a two-layer neural network that approximates this function with that precision (Hornik et al., 1990). Hence the advantage of nonlinear dendrites cannot be expressivity per se. Nevertheless, these nonlinearities can play a role under resource constraints. Investigating this hypothesis, (Wu et al., 2018) compared conventional deep neural networks to deep networks of neurons with nonlinear dendrites (see Section 3.2 for more details). Non-linear—as opposed to linear—dendrites in interneurons were also predicted to provide important resource savings in biological model networks: they enable the storage of individual associative memories via the use of significantly fewer neurons, that fire at lower rates and whose activity is sparser than networks with linear interneurons (Tzilivaki et al., 2019). A similar contribution of active dendrites in combination with synapse clustering was seen in associative fear learning. Faster learning was experimentally reported in mutant mice with increased synaptic turnover and increased synapse clustering. Computational modelling showed that increased synaptic turnover underlined the increase in synapse clustering and enhanced the sparsity of associative memories learned by model neurons. This enhancement in sparsity can minimize interference and enhance the storage/recall of multiple memories (Frank et al.,

---

**Table 1.** Structural plasticity learning rule for rewiring axonal connections on a simple two layer model of neurons with nonlinear dendrites.

- The weight of an existing synaptic contact could be increased or decreased (weight change) and an existing contact could be eliminated and replaced with a new contact from a different afferent input (structural change).
- The contacts targeted for elimination were those firing out-of-sync with other synapses on the same branch and/or the neuron. A synapse was instead stabilized, via weight increases, if it fired synchronously with other synapses on a branch and the branch response was correlated with the neuron response. The (de-) correlation was identified by maintaining a fitness score $C$ as a synaptic tag.
- A structural change was implemented via axonal re-wiring, whereby (a subset of) pre-synaptic axons were allowed to form and retract synapses randomly within the $m$ branches.

2018). Overall, these findings suggest that dendritic non-linearities enable biological networks to learn and store information in a way that saves resources.

Maybe the most important resource constraint in the brain is connectivity. While most of the brain volume is consumed by axons, the neurons in the brain are still extremely sparsely connected. Modern imaging techniques have shown that this connectivity matrix is highly dynamic, with synaptic connections being established and removed on the time scale of hours and days (Holtmaat et al., 2005; Stettler et al., 2006; Kasai et al., 2010; Loewenstein et al., 2011; Rumpel and Triesch, 2016). It was hypothesized that these dynamics are used by the brain to optimize network function under severe constraints on connectivity (Kappel et al., 2015; Kappel et al., 2018). A recent study has investigated the role of dendrites in this process (Limbacher and Legenstein, 2020), see Fig. 2C. In a two-layer neuron model with temporal dynamics of dendritic spikes, the authors conditioned synaptic plasticity on local dendritic plateau potentials. During plateau potentials, pre-synaptic activity potentiated synaptic connections, while inactive synapses were depressed. The plasticity also contained a stochastic component (Yasumatsu et al., 2008; Dvorkin and Ziv, 2016) which included the removal of weak synapses and the stochastic addition of dormant connections. Mathematically, this process can be described by a Fokker–Planck equation and it can be shown that it implements a sampling process over connectivity configurations where co-active inputs are more likely to cluster on branches. Several studies (Hussain et al., 2015; Roy and Basu, 2016) have also demonstrated that the use of neurons with nonlinear dendrites allow rewiring based rules to learn sparse connectivity matrices which consume less synaptic resources than their conventional ANN counterparts, while achieving similar classification accuracy on benchmark machine learning data sets. These are discussed in more details in Section 3.1. Such rewiring mechanisms have been proven useful also for other technical applications, for example to train sparse deep neural networks (Bellec et al., 2018) or connectivity-constrained neuromorphic hardware (Yan et al., 2019; Bhaduri et al., 2018).

### 2.3. P3: Utilization of internal learning signals

Biological plasticity rules may not necessarily match the best performing learning algorithms in ANNs. These algorithms typically rely on backprop, a method where precise error signals are backpropagated through the network, from the outputs (where errors are readily available) to neurons closer to the input. This learning principle has been considered biologically implausible (Crick, 1989). It is clear that learning-related signals are communicated within the brain (Roeper, 2013; Engelhard et al., 2019), for example in the form of neuro-modulators such as dopamine and acetylcholine. But these signals are quite different from error signals used in deep learning as they are not neuron-specific, they are temporally imprecise, and come from specialized structures. In contrast, many molecular and electric sig-

nals such as local dendritic spikes, the backpropagating action potential, $Ca^{2+}$ concentrations, and molecular messengers, are available within a neuron. Hence, if a neuron can be considered equivalent to a multi-layer neural network, there are many possibilities to orchestrate the tuning of synaptic weights and other plastic neuron parameters within that neuron-specific network.

Such biological learning rules, in conjunction with active dendrites, were shown to sub-serve important functions in biological networks. For example, Kastellakis et al. (Kastellakis et al., 2016) showed that clustered plasticity together with the plasticity of neuronal excitability underlies the linking of information across time. This linking is achieved by directing the storage of subsequent memories in common dendrites of overlapping neuronal populations, where axons from subsequent memories co-cluster within active dendrites. Similar findings were also seen with respect to the formation of associated memories: the binding of information from separate input streams was achieved via co-clustering of respective inputs from these streams within active dendrites (Kastellakis et al., 2016; Legenstein and Maass, 2011; Limbacher and Legenstein, 2020). The general idea here is that local voltages in dendritic subunits can reinforce plasticity of co-active inputs within a dendritic subunit, leading to synaptic clustering. The backpropagating action potentials on the other hand could serve as a global signal (with respect to the neuron) that can communicate to other branches that the neuron is already active for the current input. This can be utilized by plasticity rules to ensure that different branches are used to encode different associations (Legenstein and Maass, 2011; Limbacher and Legenstein, 2020).

Recently, experimental evidence was accumulated which shows that plasticity of synapses in the distal apical dendrites of cortical pyramidal cells are gated by instructive signals (Magee and Grienberger, 2020). Several such instructive signals have been identified, including $Ca^{2+}$ spikes (Kampa et al., 2006; Kampa et al., 2007), various types of dendritic plateau potentials (Gambino et al., 2014; Bittner et al., 2015; Bittner et al., 2017; Magee and Grienberger, 2020), or signals of synaptic connections from other brain areas (Doron et al., 2020; Aru et al., 2020). In a recent theoretical study, Rao et al. (Rao et al., 2021) investigated synaptic plasticity at distal dendrites in this context. They showed that with two integration compartments, one somatic and one at the dendritic tuft, synaptic plasticity gated by an instructive signal can give rise to a powerful learning method, a spike-based approximation of logistic regression. Logistic regression can inherently deal with noisy data (as opposed to other methods such as the perceptron learning rule (Moldwin and Segev, 2020)). This property is important when one assumes for example that instructive signals are provided by top-down connections from higher brain areas and that the learning goal is to predict bottom-up sensory inputs on their basis (Larkum, 2013). Predictions are inherently uncertain, leading to noisy overlapping prediction targets for which perceptron learning would fail. While the two-compartment model gives rise to a linear logistic regression model, they show that much

more complex target distributions can be learned when several dendritic compartments are considered.

The interplay between local plasticity rules and nonlinear dendritic processing has also been studied in detailed compartmental neuron models. Iannella et al. (Iannella et al., 2010) have observed the emergence of spatial synaptic clusters in simulations of detailed compartmental models of L2/3 pyramidal neurons. In another study of detailed L2/3 pyramidal neuron models, it was found that nonlinear dendritic effects give rise to a diverse set of plasticity rules along the spatial extension of the model neuron and that distal plateau potentials can switch the direction of plasticity at other synapses (Bono et al., 2017).

Finally, it has been argued that the electrotonic segregation between the basal dendrites and the apical tuft may be utilized by pyramidal cells to combine error signals—coming from feedback connections to the dendritic tuft—with feed-forward information arriving more proximal to the soma (Kording and Konig, 2001; Urbanczik and Senn, 2014; Guerguiev et al., 2017; Whittington and Bogacz, 2017; Sacramento et al., 2018; Richards and Lillicrap, 2019; Whittington and Bogacz, 2019), see Section 3.2 for more discussions on its relation to deep learning.

## 2.4. P4: Enabling continual learning

Current state-of-the-art deep neural networks can be trained on a comprehensive selection of tasks on which they achieve impressive performance (LeCun et al., 2015). However, these neural networks typically suffer from a phenomenon known as catastrophic forgetting (Robins, 1995; McCloskey and Cohen, 1989; French, 1999). When they are trained on a new task, they forget what was learned before. This phenomenon, which is caused by updates of network parameters that fail to preserve prior knowledge (Robins, 1995), makes these networks unsuitable in a continual learning setup. The brain, however, has clearly implemented a clever and efficient algorithm to continually acquire knew knowledge without forgetting the past.

A possible role of dendrites in continual learning was studied by (Cichon and Gan, 2015). Cichon and Gan showed that a synaptic plasticity rule that depends on local dendritic activity enables different inputs to be processed on different dendritic branches, reducing their interference. More specifically, they found that different motor learning tasks induce dendritic spikes in different dendritic branches of pyramidal cells (PCs) in the mouse motor cortex. A possible explanation for this finding is that synaptic inputs are clustered onto dendritic branches in a task-specific manner. Importantly, when this task-specific segregation of dendritic activity was pharmacologically disrupted, the animals could still learn the current task, but performance on an earlier learned task degraded, indicating a role of synaptic clustering in preserving previously required skills.

A recent computational study (Limbacher and Legenstein, 2020) showed that in a spiking neuron model for PCs with dendritic non-linearities, synaptic clustering can indeed serve as a mechanism to mitigate catas-

trophic forgetting in a pattern memorization task (Fig. 2D). Their model combines a generic stochastic rewiring principle (Kappel et al., 2015) with an NMDA-spike-mediated synaptic plasticity rule. They found that this synaptic rewiring model leads to synaptic clustering of temporally correlated inputs on dendritic branches. This is in line with the idea that functionally related inputs form spatial clusters on dendritic branches (Poirazi and Mel, 2001; Govindarajan et al., 2006; Kastellakis et al., 2015). Importantly, the findings by (Limbacher and Legenstein, 2020) indicate that a completely local dendritic-spike-dependent plasticity rule alone is not sufficient to protect previous memories in a sequential learning task. The authors showed that by combining their model with an additional spike-timing-dependent plasticity (STDP) mechanism, that introduces competition between dendritic branches (as proposed in (Legenstein and Maass, 2011)), not only increases the continual learning capability of their model, but that this mechanism also increases the number of input patterns that can be stored by a neuron.

An increased memory capacity was also observed in a hippocampal network model with active dendrites (Kaifosh and Losonczy, 2016). The memory-encoding scheme in this model was based on non-linear interaction between anatomically and functionally segregated inputs onto dendrites of two-compartmental PCs in areas CA3 and CA1. The authors proposed somatic burst-firing due to interacting backpropagating action potentials and dendritic spikes as a plasticity mechanism. They found that nonlinear dendrites enhance their models capacity to store and retrieve information-rich memory engrams, which do not interfere even when a large number of highly similar memories are stored.

The above findings from experimental and computational neuroscience have inspired the development of an algorithm for artificial neural networks that addresses the challenge of continual learning in supervised and reinforcement learning scenarios (Kirkpatrick et al., 2017). This algorithm, referred to as elastic weight consolidation (EWC), acts by slowing down learning of certain weights identified based on how important they are to previous tasks. The authors show that this task-specific synaptic consolidation allows multiple tasks to be learned with networks of fixed size, and where tasks with related structure reuse shared components of the network. Hence, EWC allows artificial neural networks to support continual learning of challenging tasks at large scale. A mechanistic explanation for the EWC algorithm could be based on NMDA spikes. Bono and Clopath proposed that during associative learning, an NMDA-spike-mediated plasticity mechanism in basal dendrites acts similarly to synaptic consolidation. In networks of two-compartment neuron models, this type of NMDA-spike-mediated plasticity can protect stored associations from subsequent modifications when learning new associations (Bono and Clopath, 2017). Finally, the selective consolidation of some memories, but forgetting of others, was studied by (O'Donnell and Sejnowski, 2014). The authors proposed a rule dependent on protein synthesis, which allows for selective memory consolida-

tion even for simultaneous events that are represented by the same neural population. The selectivity in their model depends on functional clustering of synapses on dendrites at the neuron level and the overlap of neuronal activity patterns at the circuit level.

## 3. APPLICATIONS: MACHINE LEARNING

The previous section 2 highlighted the computational benefits afforded by dendrites in simulated networks and pointed out four major principles. However, the surveyed literature did not *quantify* the efficiency of using nonlinear dendritic processing in real-world tasks such as image recognition. Such tasks are used to benchmark conventional machine learning algorithms and have become the defacto method for comparing new algorithms in the machine learning community. In this section, we review the literature reporting use of dendritic properties for resource efficient performance on several real-world data sets such as UCI (Dua et al., 2017), MNIST (LeCun et al., 1998), CIFAR (Krizhevsky et al., 2009) etc. The MNIST data set comprises 60,000 examples of handwritten digits from 0 to 9 for training and a separate test set comprising 10,000 samples. These samples are good for training machine learning systems aiming to perform optical character recognition. The CIFAR data set is also for visual object recognition comprising 50,000 training and 10,000 test examples for objects belonging to various categories spanning vehicles to animals. Lastly, the UCI repository has a wide variety of data sets such as medical records for prognostics (breast cancer, diabetes etc), time series data (e.g. audiology, arrhythmia etc) and many more (e.g. radar data in ionosphere data set; plant type classification in iris data set etc.). The UCI repository is popular to test the performance of ML algorithms due to its wide diversity of data spanning low to high dimensionality of input data and sample numbers. All these data sets primarily test the generalization capabilities of the algorithm on unseen data. We refer to such data as being "real-world" since they reflect the variability of data distributions faced by ML algorithms when deployed in different applications. We review work on shallow networks first (sub-section 3.1) when operating on these data sets followed by studies related to Deep Learning architectures (in sub-section 3.2), for which the community has shown a special interest. All studies reviewed in this section adhere to the principles P1 and P2 of improved expressivity and resource usage described in Section 2.

### 3.1. Shallow Learning

Numerous studies have utilized the non-linear amplification properties of dendrites along with some form of plasticity to solve real-world learning tasks. Most of the plasticity mechanisms used in such studies (Fig. 3) include a combination of structural and weight plasticity as described in Section 2.1. Beyond synaptic updates, weight plasticity is also used to change the branch coupling coefficients while delay plasticity is used to change the presumed distance of the synapse from the soma, a property that is particularly useful

when learning spike-timing coded patterns. The flexibility afforded by a network of neurons equipped with dendrites is summarized in Fig. 3 which shows the various degrees of freedom as (i) synaptic weights $W$, (ii) clustering of connections on one branch and different compartments within that branch (iii) with different delays and choice of branch coupling coefficients $\beta$ for the various branches when connecting to the soma. The function $b()$ denotes a lumped branch nonlinearity (as denoted earlier in Section 2.1) and $g()$ is the lumped somatic nonlinearity typically selected as a heaviside function to denote the all or none behavior of the action potential generation. This simplified diagram ignores temporal dynamics which can be obtained by replacing $g()$ with differential equations to represent different neuron models such as leaky integrate and fire or Hodgkin–Huxley. The simplified model can be considered as an approximation to the spiking model in the rate coding regime (Hussain et al., 2015) The following sections discuss studies based on their use of the three plasticity mechanisms mentioned above.

*3.1.1. Structural plasticity.* The effect of nonlinear dendrites and structural plasticity (for binary synaptic connections) in memory capacity was reviewed in Section 2.1. This earlier model (Poirazi and Mel, 2001) can be mathematically thought of as modifying the structure of connections within dendritic branches shown in Fig. 3 without changing delays or $\beta$. Structural plasticity can also create changes in $W$ by allowing multiple connections from the same afferent input to be created within a dendrite. The learning algorithm used in this work was described in Table 1 and was shown to achieve higher performance on a binary classification task.

However, (Poirazi and Mel, 2001) had only used this structural plasticity mechanism for pattern discrimination/memorization. Generalization to unseen data, which is important for real world machine learning problems, was not explored. Moreover, (Poirazi and Mel, 2001) only used static binary inputs which can be thought of as equivalent to a binary rate code—its application to spike-timing-coded patterns was also missing. Lastly, hardware realization of these systems on low-power neuromorphic substrates is essential for many applications (Basu et al., 2018)—the previous learning rules as described in Table 1, used terms with derivatives of dendritic and somatic nonlinearities and hence were not amenable for efficient hardware realization using integrated circuits.

Towards this goal, (Hussain et al., 2014b; Hussain et al., 2015) created a hardware friendly learning rule whereby he introduced a modification of the fitness score $C$ by replacing the gradient of $b()$ with the function itself. The change was based on the intuition that for monotonically increasing functions, this should not change the ranking. Further, the gradient of the somatic nonlinearity $g()$ was replaced by a ternary valued signal to only guide the direction of change in $C$. Lastly, to improve generalization to unseen data, a different function $g_{margin}()$ with a smoother slope (instead of the heaviside function $g()$) was used to include a margin in the training process sim-
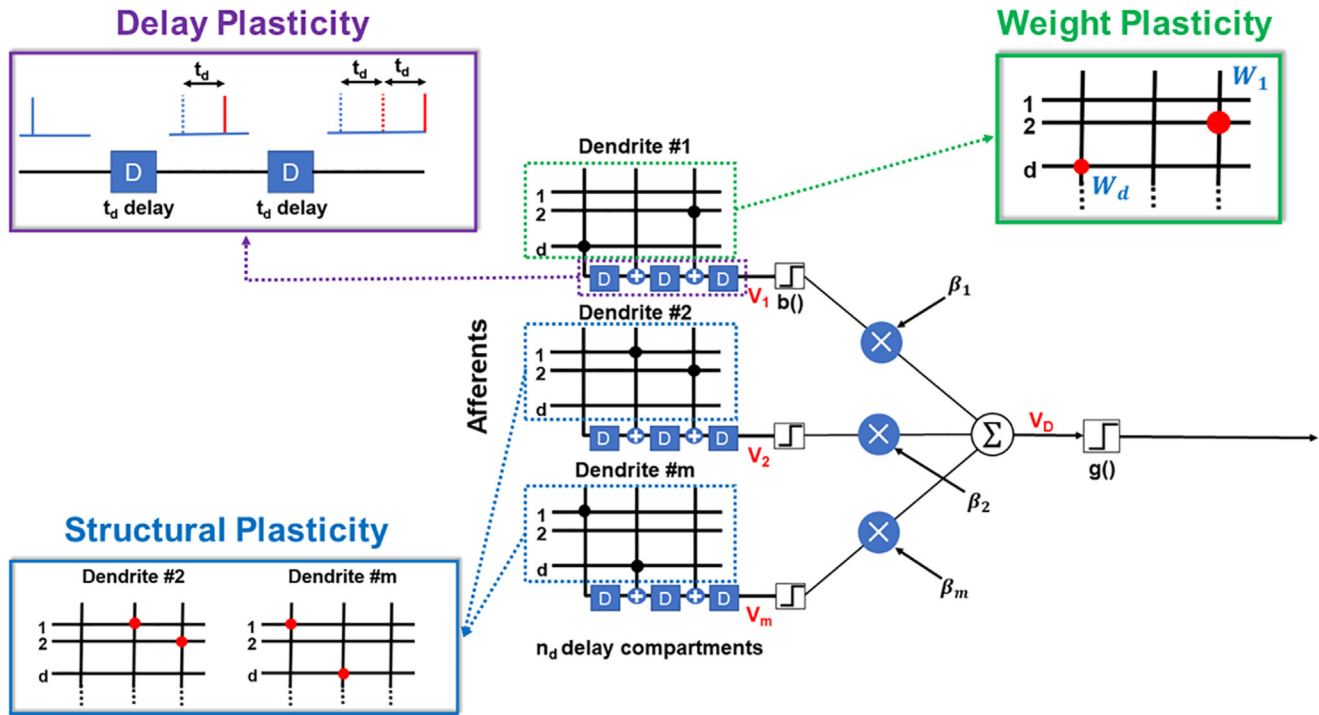
**Fig. 3.** A generic network diagram of a neuron model equipped with dendrites. The degrees of freedom that can be potentially exploited by a learning algorithm where the function $b()$ denotes a lumped branch nonlinearity and $g()$ denotes the somatic nonlinearity. Dendritic processing has been used for several learning tasks which can be categorized according to the type of plasticity mechanisms used: structural, weight or delay plasticity. The learning algorithm can change the weights of the synapses $W$ (weight plasticity) as shown by the larger circle connecting afferent 2 vs afferent $d$ on Dendrite #1. Secondly, the neuron can learn the clustering of the synapses in a compartment via structural changes or rewiring (structural plasticity). For example, afferents 1 and 2 are clustered on Dendrite #2 while afferents 1 and $d$ are clustered on Dendrite #m due to association among these respective inputs. Third, the algorithm can also learn the desired distance from soma for each afferent thus affecting delays (delay plasticity). This is particularly useful for input patterns encoded temporally. Lastly, learning can happen by changing branch coupling coefficients $\beta$ as well. This signifies the influence a particular dendritic branch voltage ($V_1$, $V_2$ etc) can exert on changing the somatic voltage.

ilar to support vector machines (SVM) in conventional machine learning. This hardware–friendly learning rule was shown to work as well as the original one and the inclusion of margin allowed the network to classify real world data sets from the UCI repository. Further, it was shown in (Bhaduri et al., 2018) that the margin based training enabled hardware realizations of these networks to be resilient against statistical variations in analog integrated circuits. While these results were obtained in a batch mode of training, (Hussain et al., 2015) also showed that a spike-based learning rule—Branch-specific Spike-time dependent structural plasticity (BSTDSP)—was capable of computing the fitness score in an online fashion. This work was later extended by the authors in (Hussain and Basu, 2016) to demonstrate classification of handwritten digits from the MNIST data set. The authors also showed that the morphology of the neurons in terms of number of dendritic branches could be modified according to the difficulty of the classification problem, leading to a better trade-off between classification accuracy and synaptic resources needed.

The use of spike-time based codes was first addressed in (Roy et al., 2013; Roy et al., 2014) where the authors used a neuron with nonlinear dendrites (NNLD) as the readout layer for a liquid state machine (LSM) (Maass et al., 2002), a type of spike based reser-voir computing architecture. The output of the "liquid" or reservoir is a spatio-temporal spike train which is the input to the NNLD. To utilize the earlier fitness score-based structural modification framework, the score $C$ was modified so that the binary input $X$ is replaced by an analog value equivalent to the post-synaptic current (PSC) at the synapse (obtained by convolving the spike train with a decaying exponential). To accelerate the learning, the derivative of $b()$ was used to compute the fitness as dictated by backpropagation. However, to make this hardware-friendly, $b()$ was chosen as a quadratic function so that the derivative was proportional to the input of the function and easily computed. It was shown that the NNLD with binary synapses outperformed the conventionally used parallel perceptron readout (PPR) in terms of classification accuracy while utilizing much fewer synaptic resources. The authors also extended this approach in (San et al., 2014; Roy et al., 2015) to classify spike latency patterns similar to the ones in (Gutig and Sompolinsky, 2006). They showed that other learning rules, such as the Tempotron (Gutig and Sompolinsky, 2006), could be used to compute the fitness score $C$. While this rule requires knowledge of the full spike pattern to calculate $C$, online variants proposed in (Gutig and Sompolinsky, 2006) could potentially be used here as well.

*3.1.2. Weight plasticity.* The papers discussed in Section 3.1.1 modified the structure of the connection matrix and could also modify $W$ to a certain extent; but they all assumed equal unit strength of branch coupling $\beta$ for each dendritic branch to the soma. Instead, Tapson et al. (Tapson et al., 2013) explored the capacity of a network where the structure of the connections is fixed, values of $W$ are random and only values of $\beta$ are learned. Since only one layer of weights are learned, it can be done in a batch using the Moore Penrose pseudo-inverse method. This is similar, in spirit, to the Neural Engineering Framework (NEF) where the input is projected to a high dimensional space using a network of randomly weighted neurons. However, in contrast to NEF where many neurons are needed to implement a function, here dendrites are used as a substitute. Moreover, the method has been shown to work for spike-time codes instead of rate codes as employed in NEF. While this work used a batch update mode, an online pseudo-inverse update method (OPIUM) may also be used as pointed out by the authors. Note that the modification of branch coupling strengths were studied from a theoretical viewpoint in (Legenstein and Maass, 2011) and this was discussed earlier in Section 2.3.

Finally, in addition to supervised classification schemes (Hussain et al., 2015; Roy et al., 2014) (whereby knowledge of the correct class for each pattern is assumed), weight plasticity was also implemented in unsupervised learning (whereby neurons alter their weights based on self-organizing principles). An important, bio-inspired learning rule in this category is spike-time-dependent plasticity (STDP)—a Hebbian rule where the weight change of a synapse depends on the time difference between the firing of pre- and post-synaptic neurons. This is usually implemented in an online fashion where pre-spikes update the weight based on a memory trace of the post-spike and vice versa. STDP has been shown to be useful in detecting repeating patterns of spikes among a large number of distractors as well as classifying spatio-temporal spike patterns. In (Roy and Basu, 2016), the authors use NNLD in a winner-take-all (WTA) architecture where the competition enables neurons to tune to different spike patterns. An STDP like Hebbian rule is used to update the fitness score; but it is modified to be branch specific by multiplying the derivative of $b()$ to the pre- and post-synaptic memory traces. This fitness score is again used to replace synapses after every batch of patterns. It was shown that the NNLDs eventually learned to specialize to different spatio-temporal patterns and their order of firing could be used to discriminate between patterns.

*3.1.3. Delay plasticity.* The studies in Section 2 as well as Sub-sections 3.1.1, 3.1.2 focused on modifying network structure and/or synaptic weights. However, depending on the location of the synapse on the dendritic tree, there will be varying delays for its input spike to affect the somatic integration. This has been exploited to learn different classes of spike latency patterns in (Hussain et al., 2014a). A neuron in this network modifies delays for each of the input spikes using

a learning rule adapted from the tempotron rule (Gutig and Sompolinsky, 2006) while weights are kept constant—hence, it is termed "Deltron". Similar to the Tempotron, the network tries to learn that part of the pattern which is most salient, i.e. has the highest number of synchronous inputs. The authors demonstrated the memory capacity of such learning rules as a function of the load (ratio of number of patterns to number of synapses) and showed classification accuracy of $\approx 80\%$ for load of 2.

A last category of research related more to a signal processing aspect of passive dendritic cables did not use plasticity directly, but relied on their analogy with continuous time hidden markov model (HMM) formulations (Hasler et al., 2007; George et al., 2013). In this line of work, the authors draw an analogy between the activation along a dendrite and a traveling wave equation that is similar to the continuous time formulation of hidden markov models (HMM) (Lazzaro et al., 1996). The traveling wave effect was found to be more prominent when the dendritic diameter was assumed to increase gradually towards the soma. The authors used this system to perform simple keyword classification where the spiking inputs to each compartment signified the probability of detecting a phoneme and the dendritic line helped identify the correct sequence. Different words were represented by separate dendritic branches and a WTA is used to select the winning dendrite. While not explicitly shown by the authors, to generalize this scheme to classify a given set of patterns, one would need to learn the location of connection of an input afferent on each dendritic branch thus selecting appropriate delays. Hence, this can be thought of as a form of delay or structural plasticity.

The studies described in Section 3.1 used only one form or another of the palsticity mechanisms described above, without combining them. Some initial work in this direction was done in (Hussain and Basu, 2016) to combine weight and delay plasticity through structural modification, but the blow-up of the parameter space prevented large scale explorations. With the availability of better computing resources thanks to the Deep Learning revolution, we envision this to be a useful avenue of future research. Further, the theoretical studies in Section 2 have rarely considered the effect of dendrites in enabling learning precise temporal codes of spikes, a topic more frequently explored in engineering applications as described in Section 3.1.3. With the rising popularity of event-driven sensors and spiking neural networks, as well as better computing resources, we envision more studies in this direction. Table 2 summarizes the different approaches according to their plasticity mechanisms as well as salient features of each of them. The next section introduces more recent work at the intersection of dendritic processing and deep neural networks.

## 3.2. Deep Learning

As shown in previous sections, dendritic compartmentalization provides an array of powerful features to neurons, the core information processing units of our brain. Moreover, both experimental and computational studies in Neuroscience are pointing to the significant advantages provided by dendrites in

**Table 2.** Comparison of the different types of dendritic learning reviewed in this section.

| Ref | Plasticity | Supervised | Online | Batch | Spike-time based | Comments |
|---|---|---|---|---|---|---|
| Hussain 2014 | Delay | Yes | No | Yes | Yes | Learning rule inspired by tempotron; Online approximation possible |
| Hussain 2015 | Structural (Weight) | Yes | Yes | Yes | No | Online update of fitness score, batch update of weight; Spike rate code; UCI data set |
| Roy 2014 | Structural (Weight) | Yes | Yes | Yes | Yes | Online update of fitness score, batch update of weight; Decode output of reservoir in Liquid state machine |
| Roy 2016 | Structural (Weight) | No | Yes | Yes | Yes | Unsupervised learning–branch specific STDP on neurons in WTA |
| Tapson 2013 | Weight | Yes | No | Yes | Yes | Inspired by NEF, first layer W is random; Online approximation possible |
| George 2013 | - | Yes | - | - | Yes | Dendrite cable used as HMM; Input to dendrite is probability of phoneme; Application of wordspotting |
| Hussain 2016 | Structural (Weight) | Yes | No | Yes | No | MNIST data set; Dendrite morphology depends on problem difficulty |
| Wu 2018 | Weight | Yes | No | Yes | No | Improved model expressivity; UCI data set |

STDP: Spike-time dependent plasticity; WTA: Winner take all; NEF: Neural Engineering Framework; HMM: Hidden Markov Model.

neuronal and circuit computations. However, in the emerging field of deep learning research, neurons are still modeled as linear summation units followed by the nonlinear output function, also known as "point neurons". The natural question that arises is whether adopting an architecture design that resembles that of dendrites would improve the performance of modern deep neural networks? We briefly discuss the few studies that explored this question (Wu et al., 2018; Jones and Kording, 2021).

To understand how furnishing artificial neurons with dendrite-like features could affect deep neural network models, we assess the following characteristics: (1) The expressive power of dendritic neural networks; (2) the ability of dendritic neural networks to learn target functions through learning algorithms, such as Stochastic Gradient Descent (SGD); (3) the generalization performance of trained dendritic neural networks and the inductive biases associated with dendritic structure under different configurations and (4) the performance of dendritic neural networks on real-world applications. Among these, the first two points are similar to the principles P1 and P2 discussed in Section 2.

It is commonly accepted that a deep network can achieve a better approximation to arbitrary functions (namely it has higher expressivity) than a shallow network, when equipped with an equivalent amount of parameters (Poggio et al., 2020; Montúfar et al., 2014). Replacing point neurons in a neural network with neurons equipped with dendritic structure is somewhat similar to inserting extra layers into a model; thus we expect neural networks fitted with dendrites to offer more fitting power. The work in (Poirazi and Mel, 2001; Poirazi et al., 2003) gives us an indication of the extra fitting power dendritic neural networks can offer on a theoretical basis and on approximating neuronal firing behavior (discussed in Section 2.1). In (Wu et al., 2018), the authors assess model complexity through empirically measuring the number of linear regions their dendritic neural networks models can map input-space into (Montúfar et al., 2014; Raghu et al., 2017). Their results show that segregating a point neuron into smaller dendritic subunits indeed leads to a much larger number of linear regions, thus implying greater expressivity.

Model expressivity describes the potential of a model to approximate certain functions of high complexity, i.e. whether there exist solutions in the parametric space of a model that can give a good approximation of a target function. However, the goal of model learning is to identify those solutions from a limited set of training examples. While it is known that such a training task is an NP-complete problem (Blum and Rivest, 1992), empirical approaches, such as the SGD algorithm, can nevertheless give outstanding parametric solutions for the problem. This also implies that the expressivity of a model might not materialize to good solutions. Hence, for neural networks equipped with dendritic components, appropriate learning algorithms must be used so as to arrive at desirable solutions. In Wu et al. (Wu et al., 2018), the authors utilize the Maxout activation function (Goodfellow et al., 2013) to mimic the active dendritic properties of biological neurons. The adoption of this activation function, presumably also allows the gradient-based optimization algorithm they used to arrive at better optimal than what can be achieved when more standard activation functions, such as ReLU (Wu et al., 2018), are utilized.

Arriving at a solution that fits the training data set well, in general, does not guarantee the model to also perform well on unseen test data. That is, a model with low training loss does not always give a low test loss or a good generalization performance. Theoretical research tells us that an over-parameterized model can overfit training data, which leads to poor prediction performance on test

data. However, empirical results from deep learning research suggest otherwise. When trained with efficient data, heavily overparameterized deep neural network models greatly outperform traditional machine learning models in a wide array of tasks. Moreover, increasing the depth and width of deep neural network models, which push models deeper into the overparameterization regime, can improve model generalization performance further. Splitting a single unit point neuron into multiple dendritic subunits can, in a way, be considered as a network with greater width and depth. Will this lead to improved generalization performance as in the typical deep neural networks? In Wu et al. (Wu et al., 2018), their dendritic neural network models are tested on both small scale image data sets and a large array of data sets from the UCI repository. In their experiments, dendritic neural network models show no improved performance on image data sets while giving improved average performance over the UCI data set. It is important to point out here that the authors in (Wu et al., 2018) use a fixed connection map between layer inputs and dendritic subunits, which greatly limits the functional space the learning algorithm can explore. We hypothesize that releasing this constraint, in a way similar to neuronal structural plasticity, can lead to an improvement in both expressive power and generalization performance of dendritic neural networks. Along the same lines, work by Jones et.al., (Jones and Kording, 2021) shows that neural network models comprised by single neurons with a binary dendritic tree can out-perform linear models and reach comparable performance of standard two-layer fully connected neural networks.

Equipped with greater computational power, can dendritic neural network models be beneficial for real-world AI tasks that heavily rely on modern neural networks? To give a couple of examples of recent real-world applications of AI, behind the widely adopted modern face recognition algorithms (Kortli et al., 2020) are typically various different kinds of convolutional neural network (LeCun et al., 1995) architectures optimized for best recognition accuracy. When we call the customer service of banks or online stores, most likely our calls will be received by some sort of droids that are equipped with voice recognition systems. Again, modern neural networks are core parts behind those systems. Typically the neural networks used in those systems are recurrent neural networks or 1-D convolutional neural networks (Tandel et al., 2020). All those neural networks are based on standard point-neurons instead of neurons equipped with dendritic structure.

The most important characteristic of dendritic neural networks is that the connection map between the network layer is sparse. The computation of a dendritic layer described in Wu et al. (Wu et al., 2018) can be performed by standard General Matrix Multiply (GEMM) computing routines on general-purpose GPUs or CPUs if we ignore connection sparsity. However, such an approach is inefficient. The difficulty we need to overcome for the computation of dendritic neural networks is the same issue modern sparse neural networks need to overcome. Wu *et al.* (Wu et al., 2018) proposed an efficiency

orientated sparse network structure that constrains the connection map with an intra-group sparsity design. Such a design can allow an efficient computing dataflow for both dendritic neural networks and standard sparse neural networks, therefore, removing a critical obstacle in the real-world application for dendritic neural networks. The proposed intra-group sparsity structure is adopted by the recently released Nvidia A100 GPU for sparse neural network acceleration (Pool, 2020), making the algorithm directly accessible for dendritic neural network acceleration.

Beyond the role of dendrites in providing neural networks with greater computational power, the deep learning community is using dendritic compartmentalization to address the challenging credit assignment problem. For a typical training task, deep neural networks are trained with the back-propagation algorithm, that is, gradients with respect to the model loss metric are calculated by propagating error signal with the chain rule of differentiation. The standard back-propagation algorithm is generally perceived as problematic for biological neuronal networks. The recent success of deep learning research has re-energized the interest in understanding the credit assignment algorithms used by our brains (Lillicrap et al., 2020). It has been argued that the electronic segregation between the basal dendrites and the apical tuft may be utilized by pyramidal cells to combine error signals—coming from feedback connections to the dendritic tuft—with feed-forward information arriving more proximal to the soma. Several computational studies have proposed that this could give rise to backpropagation-like learning processes (Kording and Konig, 2001; Urbanczik and Senn, 2014; Guerguiev et al., 2017; Whittington and Bogacz, 2017; Sacramento et al., 2018; Richards and Lillicrap, 2019; Whittington and Bogacz, 2019; Illing et al., 2020; Payeur et al., 2020; Nokland and Eidnes, 2019). We believe this is a promising area of further research for both the deep learning and computational neuroscience communities.

## 4. DISCUSSION

In this review paper, we presented a summary of past research in exploiting dendritic nonlinearities for useful computations. In the first part, we summarized the neurobiology behind dendritic information processing and reviewed models for how this processing—in conjunction with synaptic plasticity mechanisms—can give rise to improved expressivity, better use of neuronal resources, and various biologically plausible learning processes. Although the processes that give rise to human intelligence are far from being understood, the emerging picture is that dendritic processing and plasticity mechanisms that depend on a variety of cell-internal signals are a crucial component of the puzzle. Recent experimental evidence indicates that plasticity of tuft synapses of cortical pyramidal cells are gated by instructive signals (Magee and Grienberger, 2020; Gambino et al., 2014; Bittner et al., 2015; Bittner et al., 2017; Magee and Grienberger, 2020; Doron

et al., 2020). It is an open question how such signals can be utilized in deep recurrent networks and how they interact with unsupervised learning mechanisms and synaptic clustering.

A role of dendritic processing in continual learning has been proposed and investigated in models (Cichon and Gan, 2015; O'Donnell and Sejnowski, 2014; Bono and Clopath, 2017; Limbacher and Legenstein, 2020). Although these models provide a proof of concept, it is still unclear whether they scale to large networks and more complex tasks. An intriguing hypothesis is that in biological networks, continual learning is achieved through a mix of plasticity processes using dendritic signals with consolidation mechanisms and an orchestration of learning in diverse areas through top-down signals. A compelling integrative theory for this is still missing.

Most work on the computational implications of nonlinear dendritic processing focused on the two-stage model (Poirazi et al., 2003) with static nonlinearities. A recent study analyzed the nonlinear properties of dendrites of mouse layer 5 pyramidal cells in prefrontal cortex by fitting a cascade filter model (Kalmbach et al., 2017). They also found pronounced nonlinear behavior. Interestingly, their study revealed that the filter had a nonlinear component that is selective to specific input dynamics (i.e., a specific rise time of the injected input). The computational implications of such dynamical dendritic processing is an interesting topic for future research.

Next, we presented a summary of applications where neurons with non-linear dendrites are used for machine learning. These methods used mostly structural, weight or delay plasticity to solve tasks from synthetic as well as real world data sets such as MNIST, UCI etc. The most commonly used architectures are shallow with one stage of learning embedded within a two-layer network while the other layer is fixed. The general observation is that for similar task accuracy, dendritic neurons use far fewer synaptic resources than their conventional fully connected counterparts.

Finally, the most recent trend is to incorporate concepts from dendritic processing into conventional deep learning architectures to take advantage of the efficient training process. A core feature associated with dendritic deep neural networks is that the inputs to a neuron are split into different computing units—thus the number of inputs received by each computing unit is significantly reduced. Constraining the input size of a computing unit is widely adopted in modern deep neural network designs. For example, modern convolutional neural networks typically utilize filters with a small receptive field size of $3 \times 3$; this significantly reduces the number of input features a computing unit receives while imposing a beneficial inductive bias to those models. Similarly, network in network (Lin et al., 2013; Sandler et al., 2018) and group convolutional network (Zhang et al., 2017) architectures further reduce the input dimension a unit receives. While the dendritic structure is similar to those model designs in the aspect of reduced subunit input size, there is a crucial difference. In a dendritic network, the outputs of dendritic branches are only used in calculating the corresponding downstream soma outputs. Such designs avoid the explosion of the hidden layer output dimensionality, and thus can greatly reduce the computing complexity of the downstream layer. Can we achieve a better balance between generalization performance and computing complexity with the dendritic neural network? This is an important question we need to answer.

The task of putting dendritic neural networks into real-world applications is still at the very early stages. Many questions remain to be addressed: How might network parameters and hidden value quantization interact with dendritic structure? Will such dendritic networks work well with modern general-purpose computing hardware? Is there one best dendritic subunit size for all learning tasks? Should we use one fixed subunit size for one specific learning task or should the sizes of dendritic subunits be varied? While there are many questions, the addition of dendritic processing to current deep networks adds a new dimension for exploration and optimization with strong evidence from biology.

## ACKNOWLEDGMENT

## REFERENCES

Ariav G, Polsky A, Schiller J (2003) Submillisecond precision of the input-output transformation function mediated by fast sodium dendritic spikes in basal dendrites of CA1 pyramidal neurons. J Neurosci 23:7750–7758.

Aru J, Suzuki M, Larkum ME (2020) Cellular mechanisms of conscious processing. Trends Cognitive Sci.

Basu A, Acharya J, Karnik T, Liu H, Li H, Seo J, Song C (2018) Low-Power, Adaptive Neuromorphic Systems: Recent Progress and Future Directions. IEEE J Emerging Selected Topics Circuits Syst 8:6–27. https://doi.org/10.1109/JETCAS.2018.2816339.

Bellec G, Kappel D, Maass W, Legenstein R (2018) Deep rewiring: training very sparse deep networks. In: International Conference on Learning Representations (ICLR).

Beniaguev, D., Segev, I., London, M., 2020. Single Cortical Neurons as Deep Artificial Neural Networks. BioRxiv..

Bhaduri A, Banerjee A, Roy S, Kar S, Basu A (2018) Spiking neural classifier with lumped dendritic nonlinearity and binary synapses: a current mode VLSI implementation and analysis. Neural Comput 30. 723–60.

Bittner KC, Grienberger C, Vaidya SP, Milstein AD, Macklin JJ, Suh J, Tonegawa S, Magee JC (2015) Conjunctive input processing drives feature selectivity in hippocampal ca1 neurons. Nature Neurosci 18:1133–1142.

Bittner KC, Milstein AD, Grienberger C, Romani S, Magee JC (2017) Behavioral time scale synaptic plasticity underlies ca1 place fields. Science 357:1033–1036.

Blum AL, Rivest RL (1992) Training a 3-node neural network is np-complete. Neural Networks 5:117–127.

Bono J, Clopath C (2017) Modeling somatic and dendritic spike mediated plasticity at the single neuron and network level. Nature Commun 8:1–17.

Bono J, Wilmes K, Clopath C (2017) Modelling plasticity in dendrites: from single cells to networks. Current Opinion Neurobiol 46:136–141.

Cazé RD, Humphries M, Gutkin B (2013) Passive dendrites enable single neurons to compute linearly non-separable functions. PLoS Comput Biol 9 e1002867.

Cichon J, Gan WB (2015) Branch-specific dendritic ca 2+ spikes cause persistent synaptic plasticity. Nature 520:180–185.

Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P (2011) Natural language processing (almost) from scratch. J Mach Learn Res 12:2493–2537.

Crick F (1989) The recent excitement about neural networks. Nature 337:129–132.

Doron G, Shin JN, Takahashi N, Drüke M, Bocklisch C, Skenderi S, de Mont L, Toumazou M, Ledderose J, Brecht M, et al. (2020) Perirhinal input to neocortical layer 1 controls learning. Science 370.

Dua, D., Graff, C., et al., 2017. Uci machine learning repository..

Dvorkin R, Ziv NE (2016) Relative contributions of specific activity histories and spontaneous processes to size remodeling of glutamatergic synapses. PLoS Biol. 14 e1002572.

Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw ND, Tank DW, et al. (2019) Specialized coding of sensory, motor and cognitive variables in vta dopamine neurons. Nature 570:509–513.

Frank, A., Huang, S., Zhou, M., Gdalyahu, A., Kastellakis, G., Wen, X., Silva, T., Poirazi*, P., Trachtenberg, J., Silva, A., 2018. Hotspots of dendritic spine turnover facilitate learning-related clustered spine addition and network sparsity. Nat Commun. 9, 422..

French RM (1999) Catastrophic forgetting in connectionist networks. Trends cognitive Sci 3:128–135.

Gambino F, Pagès S, Kehayas V, Baptista D, Tatti R, Carleton A, Holtmaat A (2014) Sensory-evoked ltp driven by dendritic plateau potentials in vivo. Nature 515:116–119.

George S, Hasler J, Koziol S, Nease S, Ramakrishnan S (2013) Low Power Dendritic Computation for Wordspotting. J Low Power Electron Appl 3:73–98.

Gidon A, Zolnik TA, Fidzinski P, Bolduan F, Papoutsi A, Poirazi P, Holtkamp M, Vida I, Larkum ME (2020) Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science 367:83–87.

Goodfellow, I.J., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y., 2013. Maxout networks. arXiv preprint arXiv:1302.4389..

Govindarajan A, Kelleher RJ, Tonegawa S (2006) A clustered plasticity model of long-term memory engrams. Nat Rev Neurosci 7:575–583.

Guerguiev J, Lillicrap TP, Richards BA (2017) Towards deep learning with segregated dendrites. ELife 6 e22901.

Gutig R, Sompolinsky H (2006) The tempotron: a neuron that learns spike timing–based decisions. Nat Neurosci 9:420–428.

Hasler P, Kozoil S, Farquhar E, Basu A (2007) Transistor Channel Dendrites implementing HMM classifiers. In: IEEE Intl. Symp. on Circuits and Systems (ISCAS). p. 3359–3362.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: CVPR..

Hinton, G., Deng, L., et. al., D.Y., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Process. Mag. 29, 82–97..

Holtmaat AJ, Trachtenberg JT, Wilbrecht L, Shepherd GM, Zhang X, Knott GW, Svoboda K (2005) Transient and persistent dendritic spines in the neocortex in vivo. Neuron 45:279–291.

Hornik K, Stinchcombe M, White H (1990) Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks 3:551–560.

Hussain S, Basu A (2016) Morphological learning in multicompartment neuron model with binary synapses. In: IEEE Intl. Symp. on Circuits and Systems (ISCAS). p. 2527–2530.

Hussain S, Basu A (2016) Multiclass classification by adaptive network of dendritic neurons with binary synapses using structural plasticity. Front Neurosci 10:113.

Hussain S, Basu A, Wang RM, Hamilton TJ (2014a) Delay learning architectures for memory and classification. Neurocomputing 138:14–26.

Hussain S, Liu SC, Basu A (2014b) Improved margin multi-class classification using dendritic neurons with morphological learning. In: IEEE Intl. Symp. on Circuits and Systems (ISCAS). p. 2640–2643.

Hussain S, Liu SC, Basu A (2015) Hardware-amenable structural learning for spike-based pattern classification using a simple model of active dendrites. Neural Comput 27:845–897.

Iannella NL, Launey T, Tanaka S (2010) Spike timing-dependent plasticity as the origin of the formation of clustered synaptic efficacy engrams. Front Comput Neurosci 4:21.

Illing, B., Gerstner, W., Bellec, G., 2020. Towards truly local gradients with clapp: Contrastive, local and predictive plasticity. arXiv preprint arXiv:2010.08262..

Jadi MP, Behabadi BF, Poleg-Polsky A, Schiller J, Mel BW (2014) An augmented two-layer model captures nonlinear analog spatial integration effects in pyramidal neuron dendrites. Proc IEEE 102:782–798.

Jones IS, Kording KP (2021) Might a single neuron solve interesting machine learning problems through successive computations on its dendritic tree? Neural Comput 33:1554–1571.

Kaifosh P, Losonczy A (2016) Mnemonic functions for nonlinear dendritic integration in hippocampal pyramidal circuits. Neuron 90:622–634.

Kalmbach BE, Gray R, Johnston D, Cook EP (2017) Systems-based analysis of dendritic nonlinearities reveals temporal feature extraction in mouse l5 cortical neurons. J Neurophysiol 117:2188–2208.

Kampa BM, Letzkus JJ, Stuart GJ (2006) Requirement of dendritic calcium spikes for induction of spike-timing-dependent synaptic plasticity. J Physiol 574:283–290.

Kampa BM, Letzkus JJ, Stuart GJ (2007) Dendritic mechanisms controlling spike-timing-dependent synaptic plasticity. Trends Neurosci 30:456–463. https://doi.org/10.1016/j.tins.2007.06.010.

Kappel D, Habenschuss S, Legenstein R, Maass W (2015) Network plasticity as bayesian inference. PLoS Comput Biol 11 e1004485.

Kappel D, Habenschuss S, Legenstein R, Maass W (2015) Synaptic sampling: a bayesian approach to neural network plasticity and rewiring. Adv Neural Inform Process Syst 28:370–378.

Kappel D, Legenstein R, Habenschuss S, Hsieh M, Maass W (2018) A dynamic connectome supports the emergence of stable computational function of neural circuits through reward-based learning. Eneuro:5.

Kasai H, Fukuda M, Watanabe S, Hayashi-Takagi A, Noguchi J (2010) Structural dynamics of dendritic spines in memory and cognition. Trends Neurosci 33:121–129.

Kastellakis G, Cai DJ, Mednick SC, Silva AJ, Poirazi P (2015) Synaptic clustering within dendrites: an emerging theory of memory formation. Progress Neurobiol 126:19–35.

Kastellakis G, Poirazi P (2019) Synaptic clustering and memory formation. Front Molecular Neurosci 12:300.

Kastellakis G, Silva AJ, Poirazi P (2016) Linking memories across time via neuronal and dendritic overlaps in model neurons with active dendrites. Cell Reports 17:1491–1504.

Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, et al. (2017) Overcoming catastrophic forgetting in neural networks. Proc National Acad Sci 114:3521–3526.

Körding KP, König P (2001) Supervised and unsupervised learning with two sites of synaptic integration. J Comput Neurosci 11:207–215.

Kortli Y, Jridi M, Al Falou A, Atri M (2020) Face recognition systems: A survey. Sensors 20:342.

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

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems (NIPS):1106–1114.

Larkum M (2013) A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. Trends Neurosci 36:141–151. https://doi.org/10.1016/j.tins.2012.11.006.

Lazzaro J, Wawrzynek J, Lippman R (1996) A Micropower Analog VLSI HMM State Decoder for Wordspotting. Adv Neural Inform Process Syst (NeuRIPS):727–733.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444.

LeCun Y, Bengio Y, et al. (1995) Convolutional networks for images, speech, and time series. Handbook Brain Theory Neural Networks 3361.

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324.

Legenstein R, Maass W (2011) Branch-specific plasticity enables self-organization of nonlinear computation in single neurons. J Neurosci 31:10787–10802.

Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G (2020) Backpropagation and the brain. Nat Rev Neurosci 21:335–346.

Limbacher T, Legenstein R (2020) Emergence of stable synaptic clusters on dendrites through synaptic rewiring. Front Comput Neurosci 14:57.

Lin, M., Chen, Q., Yan, S., 2013. Network in network. arXiv preprint arXiv:1312.4400..

Loewenstein Y, Kuras A, Rumpel S (2011) Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo. J Neurosci 31:9481–9488.

Maass W, Natsclager T, Markram H (2002) Real-time computing without stable states: a new framework for neural computation based on perturbation. Neural Comput 14:2531–2560.

Magee JC, Grienberger C (2020) Synaptic plasticity forms and functions. Ann Rev Neurosci 43:95–117.

Major G, Larkum ME, Schiller J (2013) Active properties of neocortical pyramidal neuron dendrites. Ann Rev Neurosci 36:1–24.

McCloskey, M., Cohen, N.J., 1989. Catastrophic interference in connectionist networks: The sequential learning problem, in: Psychology of learning and motivation. Elsevier. volume 24, pp. 109–165..

Mengual UM, Wybo WA, Spierenburg LJ, Santello M, Senn W, Nevian T (2020) Efficient low-pass dendro-somatic coupling in the apical dendrite of layer 5 pyramidal neurons in the anterior cingulate cortex. J Neurosci 40:8799–8815.

Moldwin T, Segev I (2020) Perceptron learning and classification in a modeled cortical pyramidal cell. Front Comput Neurosci 14.

Montúfar, G., Pascanu, R., Cho, K., Bengio, Y., 2014. On the number of linear regions of deep neural networks. arXiv preprint arXiv:1402.1869..

Naud R, Bathellier B, Gerstner W (2014) Spike-timing prediction in cortical neurons with active dendrites. Front Comput Neurosci 8:90.

Nøkland A, Eidnes LH (2019) Training neural networks with local error signals. In: International Conference on Machine Learning, PMLR. p. 4839–4850.

O'Donnell C, Sejnowski TJ (2014) Selective memory generalization by spatial patterning of protein synthesis. Neuron 82:398–412.

Payeur A, Béïque JC, Naud R (2019) Classes of dendritic information processing. Current Opinion Neurobiol 58:78–85.

Payeur, A., Guerguiev, J., Zenke, F., Richards, B., Naud, R., 2020. Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. bioRxiv..

Pecevski D, Buesing L, Maass W (2011) Probabilistic inference in general graphical models through sampling in stochastic networks of spiking neurons. PLoS Comput Biol 7 e1002294.

Poggio T, Banburski A, Liao Q (2020) Theoretical issues in deep networks. Proc Nat Acad Sci 117:30039–30045.

Poirazi P, Brannon T, Mel BW (2003) Pyramidal neuron as two-layer neural network. Neuron 37:989–999.

Poirazi P, Mel BW (2001) Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. Neuron 29:779–796.

Pool, J., 2020. Accelerating sparsity in the nvidia ampere architecture. GTC 2020..

Raghu M, Poole B, Kleinberg J, Ganguli S, Sohl-Dickstein J (2017) On the expressive power of deep neural networks. In: International Conference on Machine Learning. p. 2847–2854.

Rao, A., Legenstein, R., Subramoney, A., Maass, W., 2021. Synaptic plasticity in distal dendrites can optimize probabilistic top-down predictions in pyramidal cells. bioRxiv..

Richards BA, Lillicrap TP (2019) Dendritic solutions to the credit assignment problem. Current Opinion Neurobiol 54:28–36.

Robins A (1995) Catastrophic forgetting, rehearsal and pseudorehearsal. Connection Sci 7:123–146.

Roeper J (2013) Dissecting the diversity of midbrain dopamine neurons. Trends Neurosci 36:336–342.

Roy S, Banerjee A, Basu A (2014) Liquid state machine with dendritically enhanced readout for low-power, neuromorphic VLSI implementations. IEEE Trans Biomedical Circuits Systems 8:681–695.

Roy S, Basu A (2016) An online unsupervised structural plasticity algorithm for spiking neural networks. IEEE Trans Neural Networks Learn Syst 28:900–910.

Roy S, Basu A, Hussain S (2013) Hardware efficient, neuromorphic dendritically enhanced readout for liquid state machines. In: IEEE Intl. Joint Conference on Neural Networks (IJCNN). p. 302–305.

Roy S, San PP, Hussain S, Wei LW, Basu A (2015) Learning spike time codes through morphological learning with binary synapses. IEEE Trans Neural Networks Learning Systems 27:1572–1577.

Rumpel S, Triesch J (2016) The dynamic connectome. Neuroforum 22:48–53.

Sacramento J, Ponte Costa R, Bengio Y, Senn W (2018) Dendritic cortical microcircuits approximate the backpropagation algorithm. Adv Neural Inform Processing Systems 31:8721–8732.

San, P.P., Hussain, S., Basu, A., 2014. Spike-timing dependent morphological learning for a neuron with nonlinear active dendrites, in: IEEE Intl. Joint Conference on Neural Networks (IJCNN), pp. 3192–96..

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) Mobilenetv 2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. p. 4510–4520.

Softky W (1994) Sub-millisecond coincidence detection in active dendritic trees. Neuroscience 58:13–41.

Spruston N (2008) Pyramidal neurons: dendritic structure and synaptic integration. Nat Rev Neurosci 9:206–221.

Stettler DD, Yamahachi H, Li W, Denk W, Gilbert CD (2006) Axons and synaptic boutons are highly dynamic in adult visual cortex. Neuron 49:877–887.

Taigman, Y., Yang, M., Ranzato, M., Wolf, L., 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification, in: CVPR..

Tandel NH, Prajapati HB, Dabhi VK (2020) Voice recognition and voice comparison using machine learning techniques: A survey. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE. p. 459–465.

Tapson J, Cohen GK, Afshar S, Stiefel KM, Buskila Y, et al. (2013) Synthesis of neural networks for spatio-temporal spike pattern recognition and processing. Front Neurosci 7:153.

Tzilivaki A, Kastellakis G, Poirazi P (2019) Challenging the point neuron dogma: Fs basket cells as 2-stage nonlinear integrators. Nature Commun 10:1–14.

Urbanczik R, Senn W (2014) Learning by the dendritic prediction of somatic spiking. Neuron 81:521–528.

Whittington JC, Bogacz R (2017) An approximation of the error backpropagation algorithm in a predictive coding network with

local hebbian synaptic plasticity. Neural computation 29:1229–1262.

Whittington JC, Bogacz R (2019) Theories of error back-propagation in the brain. Trends in cognitive sciences 23:235–250.

Wu X, Liu X, Li W, Wu Q (2018) Improved expressivity through dendritic neural networks. Adv Neural Information Processing Systems:8057–8068.

Yan Y, Kappel D, Neumärker F, Partzsch J, Vogginger B, Höppner S, Furber S, Maass W, Legenstein R, Mayr C (2019) Efficient reward-based structural plasticity on a spinnaker 2 prototype. IEEE Trans Biomedical Circuits Systems 13:579–591.

Yasumatsu N, Matsuzaki M, Miyazaki T, Noguchi J, Kasai H (2008) Principles of long-term dynamics of dendritic spines. J Neurosci 28:13592–13608.

Zhang T, Qi GJ, Xiao B, Wang J (2017) Interleaved group convolutions. In: Proceedings of the IEEE international conference on computer vision. p. 4373–4382.