

UNIVERSITÄT OSNABRÜCK

MASTERS THESIS

Do You See What I See?

Applying a Spiking Neural Network to Visual Data Classification

Author:

Piper N. POWELL

Supervisors:

Prof. Dr. Tim KIETZMANN

Dr. Farbod NEZAMI

*A thesis submitted in fulfillment of the requirements
for the degree of Masters of Cognitive Science*

in the

Institute of Cognitive Science
Universität Osnabrück

August 15, 2023

I hereby certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other university.

Signed: Piper Powell

Date: 15.08.2023

I would like to thank my thesis supervisors as well as Varun Kapoor of the Kietzmann Lab and Balkaran Singh and the rest of the NeuCube team at the Auckland University of Technology for their technical and/or thematic assistance during this project. I particularly thank the NeuCube team for making the NeuCube spiking neural network available to me for testing ahead of the public release of their code.

Contents

1	Introduction	1
1.1	The Basics: A Brief Review of Visual Perception in the Human Brain	2
1.1.1	Findings in fMRI	2
1.1.2	Findings in MEG	2
1.1.3	Findings in EEG	3
1.1.4	Time and Space: A Note on the Role of Temporal Activity in Vision	4
1.2	Mind Reading: Successes in Visual Classification	4
1.2.1	Results with fMRI	4
1.2.2	Results with EEG	5
1.3	Leveling Up: Deciphering Imagined Images	6
1.4	The Present Work: Visual Imagery as a Test Run for Visual Classification	6
1.5	The New Method: Spiking Neural Networks	7
1.6	Hypotheses	8
2	Method	8
2.1	Networks	8
2.1.1	NeuCube: The Spiking Neural Network	8
2.1.2	EEGChannelNet and LSTM: The Comparison Networks	10
2.2	Dataset	11
2.2.1	Original Dataset	11
2.2.2	Custom Dataset Versions	11
2.3	Training the Networks	12
2.3.1	Training Paradigms	12
2.3.2	Replication with EEGChannelNet and LSTM	12
2.3.3	Training	12
3	Results	14
3.1	Epoch Training Paradigm	14
3.2	5-Fold Cross Validation	15
3.3	Training Time	16
4	Discussion	18
4.1	The Validity of SNNs in Visual Classification	18
4.2	Issues with the Dataset	18
4.2.1	The Banana Man Problem	19
4.2.2	Methodological Concerns	19
4.3	Insight Into Visual Classification	20
4.4	Recommendations and Future Directions	21
4.5	Implications for Visual Imagery	21
5	Conclusion	21
6	Footnotes	22
6.1	The Role of AI-Tools in This Project	22
6.2	Availability of Data and Code	22

Do You See What I See?

Applying a Spiking Neural Network to Visual Data Classification

Abstract

Extensive work across multiple fields has aimed at understanding how the brain processes what it sees. Neural data, often in the form of fMRI, MEG, and EEG data, is collected and analyzed with a variety of techniques, and newer artificial neural network (ANN) methods have seen additional boosts in success. A body of work has developed which attempts to "read" the collected brain signals to determine what image a person is viewing or imagining. This thesis will evaluate the potential of the latest generation of ANNs, spiking neural networks (SNNs) to the former task. It concludes that particularly due to their ability to handle lower amounts of data and to achieve comparable results with less training, these networks are a viable tool to develop further for visual classification and visual imagery applications. The further exploration of models deploying combinations of temporal, spatial, and recurrent elements is also recommended for this application, both within and outside of the SNN family.

1 Introduction

The human obsession with mind reading has a long and colorful history, but while the concept has existed in the lay sphere for hundreds of years, it only begins to appear in scientific literature in the 19th century (Luckhurst, 2002). Efforts to develop devices and techniques to read other people's minds have resulted in the development, at times unintentionally, of some of the most important tools of modern neuroscience. While failing to identify the neural mechanisms of telepathy, Hans Berger pioneered the use of Electroencephalography (EEG) in humans in the 1920s, and was the first to record alpha and beta oscillations in human brain activity (Kaplan, 2011; Sanders, 2021). Since that time, the topic of "mind reading" has moved from the realm of pseudo-science and crystal ball skepticism to a fully fledged and accepted avenue of research, with some modern researchers even using the once unscientific phrase freely and explicitly in formal publications (e.g. Norman et al., 2006).

Of particular interest within this field is the "reading" of images in the brain, either the product of active visual viewing or of the imagination, resulting in hundreds of papers on the subject of understanding and deciphering the neurological signals associated with these processes. Efforts to understand how the brain processes visual inputs, and to decode the resulting neural activity, has been an active topic of research for decades, with numerous publications approaching the challenge across multiple modalities and methods. Traditionally, this has been tackled by recording neural data, typically in the form of functional magnetic resonance imagery (fMRI), magnetoencephalography (MEG), or EEG data, and then analyzing the recordings with techniques including support vector machine classifiers (SVMs) and linear discriminant, multivariate, and correlational analyses, as well as a broad range of modeling techniques. This body of research has given us a vital understanding of *how* our brain sees, and a platform from which to develop techniques to decipher *what* it is seeing and, even more recently, *what it is imagining*.

These three fields, which will here be referred to as Visual Perception, Visual Classification, and Visual Imagery respectively, are by this time well established, with the third being the youngest and as yet least explored of the group. In all three, traditional analyses and machine learning techniques have resulted in extensive gains in knowledge, and in the final two, incredible feats of "mind reading." Newer artificial neural network (ANN) approaches have already propelled the fields further, and in recent years, a new generation of network has joined the ANN family with the potential to match and in some respects even outclass these achievements - spiking neural networks (SNNs).

This thesis will first review the background of the Visual Classification field, beginning with its roots in Visual Perception, and then discuss and test the application of SNNs to this task, before finally discussing their hypothetical applications in the even more challenging Visual Imagery field.

1.1 The Basics: A Brief Review of Visual Perception in the Human Brain

Understanding how the brain sees is a topic with decades of research and hundreds of articles behind it, and while it is possible to hard engineer a visual classification method without understanding or exploring the processes that create the data it will work with, these attempts are better guided with this knowledge. A review of the basics of object and scene perception in the brain is undertaken in the following sections, as these are typically the kinds of stimuli seen in visual classification tasks.

When a person views an image, they focus on aspects of that image that are most likely to be of use or interest, such as the location of living creatures, faces, and indications of danger such as blood, injuries, or potentially harmful animals (Ramanathan et al., 2010), and they achieve this on an ultrafast, microsecond scale. In people with a healthy visual system, object recognition and categorization usually takes place in as little as 100 to 150 milliseconds after they first see an image (Auge et al., 2021; Thorpe et al., 1996). Following exposure to a stimulus, activity propagates quickly along two "streams" in the visual system, termed the dorsal and ventral streams. Both begin, as would be expected, in the primary visual cortex and then propagate forward along different pathways, with the dorsal stream continuing into the parietal cortex and the ventral stream continuing into the temporal cortex. The former is responsible for more movement-oriented and spatial perception, while the latter is associated with more recognition-oriented perception. This is often referred to informally as the *ventral 'what,' dorsal 'where'* paradigm, although it should be noted that the distinction between these streams is not entirely as clear-cut as this suggests, and a significant amount of interaction occurs between them (Hebart & Hesselmann, 2012).

Human vision is traditionally regarded as hierarchical, with earlier "lower" regions and structures along these streams being responsible for processing more basic features of stimuli, while later "higher" areas aggregate this information and process the stimulus on progressively higher levels. It should be noted that activity likely does not proceed linearly and uni-directionally from these lower areas to the higher ones, as was previously thought, with a more complex interconnection likely being required to explain humans' impressive recognition performance (Herzog & Clarke, 2014). Brain activity recorded throughout the visual system during object recognition varies based on the category of object in the image (and even the exemplar thereof) and on its basic features, both when measured with high spatial resolution methods like fMRI and with high temporal resolution methods like EEG and MEG.

1.1.1 Findings in fMRI

fMRI studies have long since identified areas in the brain which selectively respond to certain categories and characteristics of stimuli, including the well-known fusiform face area (FFA) in the fusiform gyrus (Kanwisher et al., 1997), which reacts strongly to faces, and the extrastriate body area (EBA) in the lateral occipitotemporal cortex (Downing et al., 2001), which reacts strongly to human bodies. Independent patterns of voxel activity, particularly in the ventral temporal cortex, have also been found to represent object presence and object class (Carlson et al., 2003). Further work has also revealed the existence of areas in the visual cortex, termed orientation columns, that are specifically sensitive to various orientations in stimuli (Yacoub et al., 2008). Findings like these (which barely scratch the surface of the body of research in this direction) strongly indicate a crucial spatial component to visual perception and object classification, though this has not precluded the success of more temporally balanced measures such as MEG and EEG, subfields with their own success stories, particularly with respect to determining the timing of elicited neural activity (Contini et al., 2017).

1.1.2 Findings in MEG

Many visual perception studies utilizing MEG are mixed-methods studies combining both MEG and fMRI. Cichy, Pantazis, and Oliva used this combination to explore the combined spatio-temporal dynamics of object recognition, revealing that early activity in response to an image begins, as would

be expected, in the occipital lobe around 50-60 milliseconds after stimulus onset, and then progresses quickly along both the dorsal and ventral streams (Cichy et al., 2016). Brandman and Peelen subsequently used the combination to investigate the role of scenes in object recognition, finding that pathways in the visual cortex responsible for processing elements of a scene - including the transverse occipital sulcus (TOS) and the parahippocampal place area (PPA) - assist in object recognition by feeding back into pathways responsible for recognizing objects - such as the lateral occipital area (LO) and posterior fusiform sulcus (pFs). The use of MEG in this work additionally allowed them to establish the timing of the ultimate object recognition when this feedback was necessary for it (i.e. when the object itself was not easy to identify separately) at around 320 milliseconds from first view, over three times slower than object recognition under ideal circumstances (Brandman & Pee- len, 2017).

The need for this recurrent activity was subsequently identified in cases of object occlusion as well, again with MEG (Rajaei et al., 2019). The use of MEG also aided the discovery of increased beta frequency activity in Brodmann Area 17 (the visual area of the brain) in individuals with a high capacity to visualize information, while this activity was found in Brodmann Area 44 (Broca's area) in those with weaker visualization capabilities, in a paradigm testing responses under varying verbal and visual recall tasks (Nishimura et al., 2015).

Studies using MEG as the primary or only measure have also contributed to our understanding of the visual system, and in particular the timeline of object processing. Early work with MEG pointed to the role of induced gamma band oscillations in the inferior temporal gyri, superior parietal lobules, and the right middle frontal gyrus in the integration of various object information from multiple regions in the visual system during object recognition (Gruber et al., 2008). Using MEG and model fitting, Clarke, Devereux, Randall and Tyler then later discovered a progression from low-level feature processing from 50 milliseconds to size- and position-invariant shape processing around 100 milliseconds to semantic-level processing beginning around 200 milliseconds, with the final stage being key to classification (Clarke et al., 2015).

1.1.3 Findings in EEG

Work on object processing in the visual system has been equally prevalent in the EEG community. Very early work in this field is, for example, Thorpe, Fize, and Marlot's work with participants tasked with determining if a displayed photo contained an animal. During this research, they discovered that a negative voltage difference recorded by frontal EEG electrodes was associated with not seeing the target object, a reaction present only 150ms after the image was first displayed (Thorpe et al., 1996). Later work revealed a two-component progression to object recognition, with an early reaction to basic features in an image occurring roughly 135 milliseconds post-stimulus and a later true recognition response occurring between 150 and 300 milliseconds post-stimulus (Johnson & Olshausen, 2003). A similar paradigm was also indicated for face recognition, with a quick but less thorough response beginning around 140 milliseconds post-stimulus and a second slower but more refined response taking place around 250 milliseconds post-stimulus (Barragan-Jason et al., 2015).

Recorded phase-locked gamma activity in the 20-50Hz frequency range in parietal and occipital sites around 108 milliseconds post-stimulus was subsequently found to indicate familiarity with the image being viewed (Stefanics et al., 2004), findings supported by later work with gamma-based classification of trials when participants had seen, had not seen, or were not certain they had seen the present stimulus before (Rus et al., 2017). The ability to discern familiar from unfamiliar in activity in occipital regions was further supported by later work with steady-state visual evoked potentials (SSVEPs) in response to familiar and non-familiar stimuli at different flicker rates and utilizing variable resolution electromagnetic tomography (VARETA) to localize those effects (Kaspar et al., 2010). In a mixed EEG and MEG study, changes in the P1 and M1 event-related potentials (ERPs) were later found to contain information on stimulus category as early as 108 milliseconds after initial viewing when participants had been trained on the categorization task (Kietzmann et al., 2013). A separate later paper additionally pointed to the role of the amplitude in the P2 event-related potential (ERP) in indicating the openness of a scene and whether it is manmade or natural (Harel et al., 2016).

1.1.4 Time and Space: A Note on the Role of Temporal Activity in Vision

While research in this field has emphasized the importance of spatially variant activity in object recognition and general visual function, an emphasis clear in the background covered here, another line of papers reinforces the key role played by temporal elements in that activity, down to the timing of individual action potentials (often referred to as "spikes"). While some functions in the brain may operate on rate coding, meaning that information is encoded in the frequency of spikes fired by a neuron or group of neurons in a certain period of time, the ultrafast timescale of object recognition and other functions of the visual system makes this less feasible as a basis for information encoding in this system (Auge et al., 2021; Thorpe et al., 1996). Multiple studies instead point to the importance of temporal coding in this system, with the precise timing of individual spikes potentially being key for coding information relevant to features such as saliency, line orientation, contrast, and of course precise event timing (VanRullen, 2003; VanRullen et al., 2005).

1.2 Mind Reading: Successes in Visual Classification

From this background, neuroscience has revealed that object and scene processing in the brain is not a function of any single brain region or of a linear process from lower-level feature detection to higher level semantic decision making, though a basic hierarchical structure is well supported. Rather, the visual system and the brain's capacity to use it to classify objects in its environment are a complex system of interconnected regions with different areas and structures largely (though not exclusively) specialized for processing specific features or integrations of those features at specific levels, many of which feed back into each other to pass useful information between levels and assist in recognition and classification particularly under non-ideal viewing conditions.

Building off of (and supplementing) this knowledge, another line of research takes it a step further and attempts to read out the exact categories and even the exact images being viewed. While work in this vein with MEG is comparatively rare, it has been explored extensively with fMRI and, more recently, with EEG. With a massive block of research pointing to the capabilities of fMRI to reflect object distinctions based on spatial dynamics, it is no surprise that many of the most impressive advances in this field have come from work with this kind of data, but particularly in the 2010s and early 2020s, a newer block of work has also pointed to the potential of EEG in this task.

1.2.1 Results with fMRI

Perhaps the most famous early work in visual classification with fMRI data came out of the Gallant Lab at UC Berkeley in the 2000s and early 2010s. In 2008, the team developed receptive field models for each recorded fMRI voxel based on their response to natural images. These models were then fed a set of novel images to determine their theorized response to them, which was then compared with the participant's actual data to make a prediction on which of the novel images was currently being viewed. When tested with 120 novel images two months after the initial recording, the method successfully identified 82% of the images, reaching 100% when tested on a separate set of 12 novel images over a year post-experiment (Kay et al., 2008). In the same year, another team successfully utilized an SVM decoding method with fMRI data from the lateral occipital cortex (LOC) to not only determine which of two categories of object a participant was viewing, but which specific exemplar of that category they were looking at, with an average accuracy (across the different regions of interest used) of 62% across category and 55% within category (Eger et al., 2008).

Although more traditional machine learning approaches have been favored in this area, with linear SVMs and linear discriminate analyses found to be particularly useful in visual classification (Misaki et al., 2010), more recent work has also found success with neural network approaches. In a multi-category paradigm, a bidirectional recurrent network was found to exhibit an accuracy gain over traditional methods of more than 5% (Qiao et al., 2019).

Work with fMRI frequently moves beyond basic classification as well, with many projects seeking to actually reconstruct the image being viewed. In 2008, Miyawaki et al. reported the successful application of a multi-scale decoding technique to the reconstruction of images of geometric shapes from fMRI data recorded while subjects viewed those shapes, with the reconstructions able to predict the correct target image with nearly perfect accuracy (Miyawaki et al., 2008). A year after their

fMRI classification publication, the Gallant Lab team also reported the successful use of a Bayesian decoder to construct spatially and semantically accurate approximations of the images being seen by using a large bank of natural images as a prior, which while not technically a full reconstruction of exactly the image being seen, was nevertheless a significant feat (Naselaris et al., 2009).

As with classification, ANN approaches have brought about further gains in reconstruction as well in more recent years. In 2020, Mozafari, Reddy, and VanRullen reported the successful use of a generative adversarial network (GAN) to reconstruct images from fMRI data, using the GAN to develop a latent vector space from known images and fMRI responses from which novel images from novel categories could be reconstructed with an impressive accuracy of 84% (Mozafari et al., 2020). ANNs have also been proposed as possible solutions to a key problem in the fMRI reconstruction field - the need for labelled data that is relatively scarce. This has included, for example, a combination of encoding and decoding models proposed in 2019 which can allow training to be extended to fMRI data for which no image label is available and to images for which no fMRI data is available (Belyi et al., 2019). Generative adversarial networks have also been used to tackle the limited data issue, with a successful project out of Japan demonstrating the ability of GANs to allow deep networks to train and learn fMRI-image relationships for the purposes of reconstruction without the need for data-saving measures such as the use of pre-trained weights manually building in the hierarchical visual representation structure seen in the human visual system (Shen et al., 2019).

1.2.2 Results with EEG

With years of research pointing strongly to the role of spatial patterns in neural activity being key to object classification, the idea of pursuing this task with EEG, which is balanced away from spatial resolution and towards temporal, may seem strange. The low signal-to-noise ratio in EEG data also presents a specific challenge to accurate classification with visual data (Bagchi & Bathula, 2022), but the fact that this task can indeed be achieved with EEG is now a well established fact in the literature.

Interest in the use of EEG in this task has been substantially fueled by the explosion of the brain-computer interface (BCI) field, particularly in the last 10 years. fMRI, while the foundation for numerous gains in neuroscientific understanding and in neural decoding, is not a practical foundation for a widespread consumer product, involving as it does expensive and immobile technology that is not easily accessed by the majority of the population (Alazrai et al., 2020). In contrast, EEG is lightweight and highly portable, presenting an ideal platform for BCI development that has become extremely popular in the field (Alazrai et al., 2020; Anupama et al., 2014).

While BCIs using voluntary production of slow cortical potentials in live EEG to progressively select letters, often called "spellers," had already been around since the late 1990s (Birbaumer et al., 1999), the use of object recognition paradigms in BCIs wasn't introduced until the 2010s. Early work in this field by Anupama, Cauvery, and Lingaraju successfully classified viewed objects based on recorded EEG in 2014, achieving accuracies of 70%, 82% and 63% with a K Nearest Neighbors classifier, a decision tree, and an SVM respectively (Anupama et al., 2014). Additional work with SVMs working with cortical connectivity patterns demonstrated a classification accuracy of up to 80.15% in a 12-class task (Tafreshi et al., 2019). As with fMRI, visual classification EEG data has also been extensively pursued with a variety of ANN approaches.

An early application of an ANN to this task came in 2017, when Hramov et. al. reported 95% accuracy with the use of a multilayer perceptron (another name for an ANN) in a binary classification task between two views of the Necker Cube, notably creating a model which could generalize across subjects, further supporting the presence of universal processes in object recognition in the brain (Hramov et al., 2017). Successful classification has also been achieved with convolutional neural networks (e.g. Bagchi & Bathula, 2021; Kalafatovich et al., 2020; Lee et al., 2020; Palazzo et al., 2021), long short term memory models (LSTMs) (e.g. Spampinato et al., 2017; Zheng & Chen, 2021), transformer models (e.g. Bagchi & Bathula, 2022), and GANs (e.g. Mishra et al., 2023), with accuracy generally decreasing with the addition of further classes.

Although detecting inter-class differences in EEG is most typically done by aggregating numerous trials and the data from multiple participants, success has also been achieved in distinguishing

between categories on the single-trial level (e.g. Stewart et al., 2014; Wang et al., 2011), with various dimensionality reduction methods proposed as one way to account for the issues of computational cost and noise when working with single-trial data spanning all recording sites (Yavandhasani & Ghaderi, 2021).

1.3 Leveling Up: Deciphering Imagined Images

The even younger field of Visual Imagery¹ aims to take neurological data decoding a step further by decoding signals recorded not when participants were *viewing* an image but when they were *imagining* it. Although a young field with a smaller corpus than Visual Classification or Visual Perception, it has nevertheless already become a field in its own right, with successful attempts being made particularly with EEG data.

While visual imagery has been studied in fMRI studies, these studies do not typically rise to the level of "mind reading" seen in fMRI visual classification studies, but rather aim at a basic understanding of visual imagery in the brain. Key findings from fMRI studies on this topic include the dual change in activity seen in participants performing visual imagery. Posterior regions including the fusiform gyrus, posterior cingulate, and parahippocampal gyri exhibit increased activity while frontal regions including the anterior cingulate cortex, inferior frontal gyrus, insula, and auditory cortex, in addition to the early visual cortices, exhibit decreased activity, with the degree of increase and decrease correlating with the vividness of the imagery produced (Fulford et al., 2018; Zeman et al., 2010). Work by Dijkstra et. al. also revealed variations in the precuneus, right parietal cortex, medial frontal cortex, and areas of the early visual cortex based on the level of imagery vividness (Dijkstra et al., 2017).

As in Visual Classification, MEG is also not a prevalent method in Visual Imagery, and particularly as regards the classification of visual imagery data. It's lower prevalence in both Visual Classification and the classification branch of Visual Imagery may be due to it having the same portability and accessibility issues as fMRI, making it less attractive to the BCI researchers heavily influencing these fields, while also being less established than fMRI in these applications, though MEG has previously been used for the adjacent field of motor imagery (Halme & Parkkonen, 2016; Kurkin et al., 2020). EEG, however, has already been tested in a number of studies directly in the field of Visual Imagery, likely again due to the strong influx of BCI-oriented projects.

A landmark study on visual imagery with EEG came out in 2018, in which Kosmyna, Lindgren, and Lécuyer tested the then-novel idea of operating a BCI with a user imagining images as the control mechanism, versus imagining motor movements as had previously been done in imagery-based BCIs (Kosmyna et al., 2018). A flurry of papers further investigating this possibility and the EEG correlates of visual imagery has since followed (e.g. Alazrai et al., 2020; Bang et al., 2021; Lee et al., 2019, 2020; Shatek et al., 2019), with very recent work beginning to explore the reconstruction of images from EEG data as was previously done with fMRI (e.g. Shimizu & Srinivasan, 2022; Singh et al., 2023).

1.4 The Present Work: Visual Imagery as a Test Run for Visual Classification

The original target of this work was the decoding of EEG related to visual imagery, but due to the lack of a suitable dataset for this purpose and time limits preventing the creation of one, the work instead directly focuses on the application of SNNs to visual classification, and then discusses the possible utility of these networks in visual imagery based on those findings. It is important to note, however, that while neural activity during visual classification does bear some similarities to neural activity in visual imagery, these are not the same task.

Work by Kosslyn et. al. in the late 90s demonstrated activity in the primary visual cortex both when participants viewed an image and when they imagined one. Further, when they introduced

¹It is key to note that when discussing visual imagery, we are not discussing free imagination, and the images being read from the data are not novel creations, but rather recollections of previously presented stimuli. Visual imagery is explicitly defined as recalling an image from memory (Knauff et al., 2000), not generating a novel image, though there are studies whose paradigms have allowed for some degree of free-form imagination, such as when participants' imagery is cued with a verbal description (Mellet et al., 1998).

transcranial magnetic stimulation (TMS) targeting area 17 in the medial occipital cortex, it disrupted participants' ability to perceive and to imagine the images (Kosslyn et al., 1999). Research in the following year additionally found that 88% of neurons in the hippocampus, amygdala, enthorinal cortex and parahippocampal gyrus that altered their firing rate when actively viewing an image also altered their firing rate in an identical manner when imagining that image (Kreiman et al., 2000). It is also possible to predict fMRI responses to objects being actively viewed based on activity from when objects of this category are being imagined, suggesting that object recognition and visual imagery utilize some of the same structures and exhibit some of the same activity patterns (Cichy et al., 2012). The fMRI correlates of dream content (which can be taken as a form of imagery) also mimic those of the same elements when they are viewed while a person is awake (Horikawa et al., 2013). Other fMRI research also supports an overlap between these two systems (e.g. Dijkstra & Fleming, 2021), and evidence from EEG studies also suggests similar activity in the alpha band when a stimulus is viewed versus imagined (Xie et al., 2020). The similarities between perception and imagery are even strong enough that the Gallant team was able to construct an internet image search based on fMRI patterns of participants imagining works of art, which was based on a model trained with their fMRI patterns when viewing those works (Naselaris et al., 2015).

But this body of research pointing to an overlap between the systems does not mean they are identical. Particularly when discussing lower level areas of the visual system, object recognition and visual imagery can result in significantly different activation patterns. Early research suggested that visual imagery uses only a subset of the pathways engaged by perception, such as the parieto-occipital and temporo-occipital visual association areas (Roland & Gulyas, 1994), and that imagery can also activate areas responsible for other functions including language, memory, and movement based on the exact nature of the task (Mellet et al., 1998). Later research by Mellet et. al. in 2000 indicated that the primary visual cortex may not be involved in imagery at all (Mellet et al., 2000). Imagery may instead rely on a broad network of subsystems and the higher visual areas (Knauff et al., 2000). The lack of involvement of the early visual cortex and lower vision areas is disputed by other work, including the 2015 Gallant study, which found that low level features are encoded in both perception and imagery (Naselaris et al., 2015), though other studies have indicated that the receptive fields of the lower vision areas may still be different between the two tasks, being potentially larger and displaying distinct behavior in imagery (Breedlove et al., 2020).

Due to the overlap between the two systems, success in classifying visual classification data is a strong motivation for exploring a method with visual imagery data, though the differences between these systems mean that success on the latter cannot be simply inferred from success on the former.

1.5 The New Method: Spiking Neural Networks

The application of spiking neural networks to EEG classification tasks is not new, with many studies pursuing SNNs in this application because of inherent qualities of these networks which make them theoretically highly suited to such tasks.

Spiking neural networks operate on a spiking principle, with input, output, and intermediate activity in these networks all taking the form of spike trains, which are analogous to series of action potentials between real neurons (Tan et al., 2020). They are the latest generation of the neural network family and also the one which most closely mimics the function of the real human brain, with individual neurons in an SNN also typically being models of real neuron behavior, such as *Integrate and Fire (IF)* and *Leaky Integrate and Fire (LIF)* models (Gerstner, 2001; Paugam-Moisy & Bohte, 2012; Tan et al., 2020). While previous generation neural networks typically train with backpropagation, a process in which the connection weights between neurons are updated based on the discrepancy between the network's final output and the correct output, SNNs usually follow a more brain-like method for updating connection weights, often spike-timing-dependent plasticity (STDP), in which a connection is strengthened if another neuron fires shortly before the current one and weakened if it fires afterward (Paugam-Moisy & Bohte, 2012).

SNNs present a number of possible advantages when working with visual system data, and particularly with EEG data. SNNs, since they work with spike trains, are well-suited to working with temporal elements in data, for which EEG data is biased. This ability also makes them espe-

cially attractive for visual data, since the spike-based structure of SNNs is ideal for working with the spike timing element theorized to be a key component of the visual system. SNNs have also demonstrated the ability to achieve comparable results even with limited data (Taylor et al., 2014). EEG additionally presents as a form of spike train, which allows SNNs to take this form of data in a nearly raw form with less pre-processing, allowing these networks to take advantage of more information present in the original data that can be lost in manual feature selection, which many classification studies rely heavily on (Alazrai et al., 2020; Behrenbeck et al., 2019; Taylor et al., 2014).

SNNs are admittedly not without their drawbacks. There is as yet a lack of consensus on the best way to train them, and they tend to have numerous hyperparameters that still require better understanding and standardized methods for tuning (Paugam-Moisy & Bohte, 2012). The encoding of data into the spike trains they work with is also not an arbitrary task, particularly with data that does not naturally present in a form similar to a spike train (such as images) (Auge et al., 2021). Despite these challenges, however, they have already shown promise in numerous applications, including prior work with EEG, in which they have been successfully applied to tasks ranging from emotion classification to motor movement imagery (Luo et al., 2020; Virgilio et al., 2020). Zhang et al. also applied them successfully to the encoding and reconstructing of images with the visual system of salamanders (Zhang et al., 2020). Though they have not been previously applied to human visual data, their previous performance with EEG data in general, and their theoretical advantages specifically as concerns visual processes, make them a promising method to consider here.

1.6 Hypotheses

Based on the above background, this work included the following hypotheses to be tested in the paradigm detailed in the next section:

- Because they are naturally well suited for tasks requiring processing of temporal elements (Paugam-Moisy & Bohte, 2012), a potentially key element of visual data (e.g. VanRullen, 2003; VanRullen et al., 2005), and have demonstrated the ability to perform well even when fed data in a more raw form (Taylor et al., 2014), the SNN tested in this work is expected to outperform the comparison networks on a near-raw dataset.
- Because they have demonstrated the ability to perform well even when fed a lower quantity of data (Taylor et al., 2014), the SNN tested in this work is expected to outperform the comparison networks on a smaller subset of the dataset.

2 Method

All code files used in this work can be found in the GitHub repository linked in Section 6.2, with all files concerning data located in the `Data Preparation Files` directory and the master training files located in the `Master Training Files` directory. The `.idea` directory in the repository contains files needed for the IDE used for editing the code for this project that are not relevant to the content of the report. The `old` directory contains older versions of scripts and other supplemental files generated during the project that are likewise not relevant to the content or to replicating the results of the present work. An additional result aggregation file, `Final_Report.py`, is also included.

2.1 Networks

All networks tested in this work were trained on NVIDIA H100 GPUs in sessions allocated 75GB of memory and 10 CPUs. These GPUs are the property of the Kietzmann Lab and were accessed and operated via the Universität Osnabrück's High Performance Computing (HPC) network, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 456666331.

2.1.1 NeuCube: The Spiking Neural Network

For this work, the SNN chosen to be tested was the NeuCube, an SNN with a strong track record in a variety of tasks including the classification of motor imagery EEG (Behrenbeck et al., 2019), EMG data for prosthetic control (Peng et al., 2015), EEG data related to music and noise perception (N.

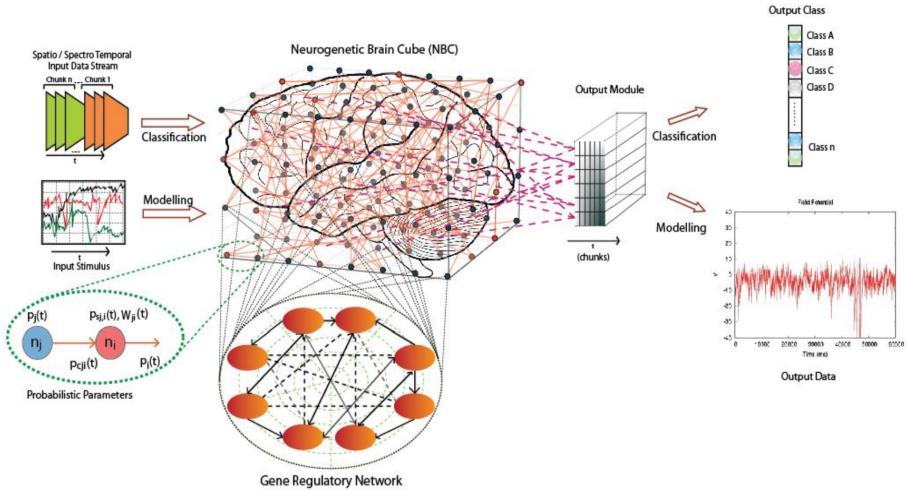


Figure 1: The NeuCube architecture. Note that the gene regulatory network is not utilized in the version of NeuCube tested in this report. Figure reproduced from N. Kasabov, 2012 in compliance with the rights and permissions for that publication, Springer Nature, under obtained license 5610141390056.

Kasabov et al., 2013), and even the prediction of ecological data (Tu et al., 2014). It has also been successfully used to classify fMRI data recorded during the reading of sentences with an affirmative versus a negative polarity, and when a participant read a sentence or viewed an image (Doborjeh et al., 2014; N. K. Kasabov et al., 2016). The SNN field has exploded in recent years, with numerous models proposed and as yet no clear agreement on precise architecture. NeuCube was therefore selected because its track record and the numerous publications applying the model to various tasks have established it as a solid representative of the broader SNN family.

NeuCube is a three-part network consisting of an encoding module, a spiking reservoir, and an output classifier (Behrenbeck et al., 2019; N. Kasabov, 2012; N. K. Kasabov, 2014). An overview of this network can be seen in Figure 1 (note that the gene regulatory element is not implemented in the version of NeuCube utilized in this work).

The input module converts the input data (in this case EEG data) into spike trains which the reservoir then takes as input. To convert input signals to spike trains, a variety of conversion algorithms have been proposed, a classic being Bens Spiker Algorithm, which uses a finite impulse response (FIR) filter in combination with thresholding and an error metric approach to determine which time points receive spikes (Schrauwen & Van Campenhout, 2003). The implementation of NeuCube tested in this work utilizes a simple delta encoding scheme, in which a threshold (delta) is set (by default 0.1) and then a one is replaced at every position in a PyTorch zeros tensor at which the difference in the activation in the sample at that point from the activation at the previous point is above the delta threshold.

The NeuCube reservoir, for which the model is named, takes the shape of a cube with a standard of 1471 individual spiking LIF neurons arranged in a brain-like pattern following the Talairich Tournoux (TT) atlas, often referred to simply as the Talairich atlas, with each neuron representing 1 cm^3 of the brain (Behrenbeck et al., 2019; N. K. Kasabov, 2014; Nowinski, 2005; Talairach et al., 1988). The reservoir is initially loaded in a small-world connectome, meaning that neurons that are closer to each other have a higher probability of being connected. Input data is then fed into the cube reservoir at the approximate locations it was collected, i.e. EEG data recorded from the F7 electrode would be input to the neurons in the reservoir cube in the section corresponding to the frontal lobe closest to the location of F7 as placed in the standard 10-10 system, a newer version of the 10-20 system which governs electrode placement in 64-256 electrode systems (N. K. Kasabov, 2014; Koessler et al., 2009).

The neurons in the reservoir are then allowed to train unsupervised, updating their connections in an STDP paradigm. While defaults for input positions (with EEG, based off of which electrode the data is from) are provided, it is also possible to enter the exact positions of recorded data. In this work, the default was used as the exact original electrode positions were not available, though assumed to have followed the standard 10-10 layout for a 128 electrode system. Because input data is fed into the reservoir at a location consistent with where it was recorded, this structure gives the NeuCube the additional ability to process spatial as well as temporal elements in the data (Behrenbeck et al., 2019; N. Kasabov, 2012; N. Kasabov et al., 2013; N. K. Kasabov, 2014).

The output classifier consists of a single output neuron for each sample in the data, which are then connected to all of the neurons in the reservoir. The output neuron is then assigned a value for that sample, which is determined by the sampling method used, and also receives the correct label for that sample. In this work, the basic Spike Count sampler was used, which simply counts the number of spikes elicited by that sample. A classifier of choice then undergoes supervised learning on the sample value - label pairs, in the case of this work a logistic regression classifier utilizing the liblinear solver. New output neurons for new samples are then assigned a label by the trained classifier (Behrenbeck et al., 2019; Doborjeh et al., 2014; N. K. Kasabov et al., 2016).

Many of NeuCube's functions in the implementation used for this work are implemented in PyTorch. Prior to training, the spiking reservoir is initialized and then, along with the chosen sampler and classifier, used to construct a PyTorch training pipeline that is then called on the data to complete the unsupervised reservoir training and the supervised classifier training, and then to predict labels for new samples.

The code repository for this project contains the custom master training file for NeuCube - NC.py. This file is used to train NeuCube in both the epoch learning and cross validation paradigms. The basic training code and the code for generating the confusion matrices is largely taken from the original NeuCube training code, with modifications made to adapt the training to the dataset used in this work and to allow it to train in both training paradigms in an identical manner to that in which the comparison networks train in those paradigms. Loss and accuracy graphing were also added.

The master training file references files for the creation of the reservoir, sampler, encoder, and other elements of the network. At this time, these files are not available to the public, but upon their release, can be added to the same environment as the provided NeuCube training file to replicate the testing with this network that is reported here.

2.1.2 EEGChannelNet and LSTM: The Comparison Networks

Two networks previously used with the dataset were used to provide a benchmark for NeuCube's performance - EEGChannelNet (Palazzo et al., 2021) and a long short-term memory (LSTM) network (Spampinato et al., 2017).

EEGChannelNet is a convolutional neural network consisting of a temporal block, a spatial block, and a residual block. In the temporal block, 1D convolutional filters filter the data *along* the channels, i.e. reading across the time dimension. The subsequent spatial block then reads the data *across* the channels, i.e. reading across location. The activity then passes through a residual block consisting of residual layers and finally a fully connected layer of size 1000, the output of which is then the EEG embedding. An additional softmax layer is then appended when employing the network on a classification task, in this case with size 40 (for the 40 image classes in the dataset) (Palazzo et al., 2021). A graphic representation of the network can be found in Figure 2.

The LSTM is a simple version of this class of network with only 1 LSTM layer in the published code (Spampinato et al., 2017). 3 versions of this network are tested in this work - the published version with 1 LSTM layer, a version with 5 LSTM layers, and a version with 10 LSTM layers. Preliminary testing with this network suggested that while increasing the layers from 1 to 5 was associated with performance improvements, 10 layers often resulted in decreased performance so depths beyond this did not warrant exploration.

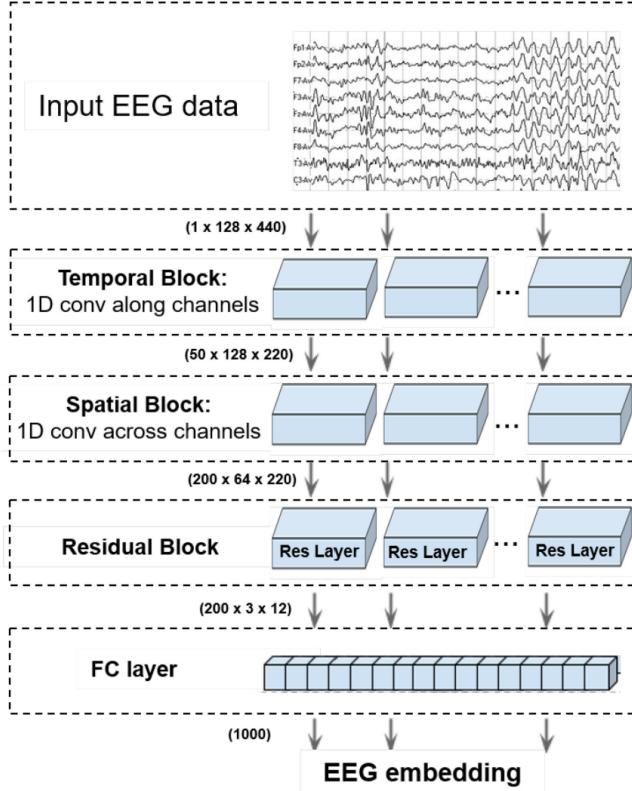


Figure 2: The EEGChannelNet architecture ©2021 IEEE. Reproduced here from Palazzo et al., 2021 in compliance with the copyright guidelines of that publication.

The code repository for this project contains the custom master training file for the comparison networks - CM.sh. This file is used for training EEGChannelNet and all three versions of the LSTM models in both the epoch learning and cross validation paradigms. The basic code for loading and applying the networks to the data is largely taken from the original training code for the comparison models, as provided in the author's GitHub repository, with modifications made to allow the comparison networks to train on the HPC more efficiently (ex. removal of the progress display) and to train in the epoch learning and cross validation paradigms in a manner identical to that in which NeuCube trains in these paradigms. Loss and accuracy graphing were also added.

The master training file references files defining the architecture of the comparison models which are not present in the repository for this project. These files can be found in the public GitHub repository of the authors for those respective papers - Palazzo et al., 2021 and Spampinato et al., 2017 - at this link.

Two additional copies of the LSTM architecture file - lstm.py - found in this repository were created to create the 5- and 10-layer LSTM models respectively, by modifying the lstm_layers variable in that file. These additional versions are not included in the repository or data entry for this work, but can be easily recreated by downloading the original lstm.py file and changing the lstm_layers variable to 5 and 10 respectively.

2.2 Dataset

2.2.1 Original Dataset

The original dataset comes from Palazzo et al., 2021 and consists of EEG data recorded while 6 participants viewed 50 images for each of 40 classes taken from the ImageNet dataset.

All images for a class were shown together, with each image shown to the participant for 0.5

seconds. A black screen was shown to participants for 10 seconds in between classes to prevent cross-class interference in the data. Following collection, the data was subsequently split into 12,000 segments of EEG data, with each individual segment containing the EEG data for one image of one class as viewed by one participant. Poor data quality subsequently resulted in the removal of 36 segments, yielding a total of 11,964 segments of EEG data recorded with 128 channels at a 1kHz sampling rate, represented in the dataset as 128 x L matrices (L being the number of samples contained in the segment, approximately 500 for each segment).

The dataset was made publicly available at the link contained in the author's GitHub repository, and is available in four different formats - the raw data, and one version each for the data bandpass filtered between 5 and 95Hz, between 55 and 95Hz, and between 14 and 70Hz.

The data is provided in the PyTorch .pth format in the form of a dictionary providing the class, image, subject, and EEG data for each segment. One .pth file is provided for each version of the dataset, and an additional .pth file contains the indices for the training, testing, and validation sets.

The custom Python file pth_info.py in the repository for this project takes an original dataset .pth file and outputs an overview of its structure and an example entry, while also checking the balance of the data across class and subject. This confirmed that the data present in the data files conformed to the description of the authors and was balanced, with data from each subject and each class making up approximately 16.68% and 2.51% of the full data respectively.

2.2.2 Custom Dataset Versions

As NeuCube takes data in the form of .csv files containing the data in a matrix with dimensions *timepoints* x *number of features* (here channels), with a separate annotation file containing the classes for each segment, the data contained in the original .pth files was processed with the create_nc_csv.py script into this form for each dataset version.

In order to allow testing of the hypothesis that a spiking neural network (NeuCube in this case) can reach a competitive level of accuracy with less data than other network types, four additional datasets were created which each contain a quarter of the data for each of the original ranges and the raw data. The quarter_data.py file was used to achieve this by randomly selecting a quarter of the data in an original .pth file, maintaining balance across participants and across classes, and then outputting the new .pth indexing file needed by the comparison networks. The quartered splits were then used to create the .csv data files and annotation files needed by NeuCube for the new, smaller datasets. Note that in the case of the comparison networks, the new data files for the quartered datasets are not actually used, but rather the new index file pulls the appropriate entries from the full dataset files during training.

2.3 Training the Networks

2.3.1 Training Paradigms

All five networks (NeuCube, EEGChannelNet, and the three versions of the LSTM) were applied to the data in two training paradigms - the epoch-style paradigm used in the original EEGChannelNet and LSTM code, and a 5-fold cross validation paradigm used in the original NeuCube code. Due to time constraints, the networks were only tested on the dataset filtered in the 5 to 95Hz range in the present work, although the other datasets and the capability to test them in all training code remain available. The decision to use the 5 to 95Hz dataset instead of the fully raw dataset was due to a correction the authors of the dataset applied to all datasets except for the raw dataset, which corrected a filtering error which resulted in erroneously high accuracies for the LSTM model in Spampinato et al., 2017.

To enable EEGChannelNet and the LSTM to train under the 5-fold cross validation paradigm, custom training code was written for these networks which prevents them from training between folds (by simply reloading the model fresh for each new fold). A new .pth splits file was also created for the 5K training and testing splits, which was generated by the create_crossval_splits.py and

`create_crossval_splits_quarter.py` files by creating the 5K split indices with the original NeuCube code and then converting these to a .pth file usable by the EEGChannelNet and LSTM training code. As the original NeuCube code generates new folds on each running of the code, these files also save a copy of the splits as Python pickle files, which are then loaded in the NeuCube code when training under the 5K paradigm.

To enable NeuCube to train under the epoch training paradigm, the pipeline is called and fit to the data on each epoch. The training, testing, and validation splits used by EEGChannelNet and the LSTM during training were converted to .csv files in the file `convert_splits.py`. When training under the epoch learning paradigm, these files are then loaded by the NeuCube training script and used to guide its construction and splitting of the Panda dataframe it constructs to contain the data.

2.3.2 Replication with EEGChannelNet and LSTM

Prior to the work with NeuCube, the performance of EEGChannelNet and the LSTM was replicated on the datasets they were tested on in their original papers (the 55-95Hz dataset in the case of EEGChannelNet and the 14-70Hz dataset for the LSTM model). The tests with EEGChannelNet generated accuracies which approximated the reported accuracy of 50%. With the LSTM, however, the reported performance of 80% was not replicated. This subsequently lead to the discovery of a critique paper noting that the released datasets were not properly filtered as described in the LSTM paper. A responding publication by the original authors later reported that, with the error corrected, the true performance of the LSTM was closer to the roughly 30% figure seen in the replication tests performed ahead of the work reported here. Further methodological questions were raised by the critiquing paper, which are detailed here in Section 4.2.2, but based on the reasoning laid out in that section, the work continued with this dataset.

The performance of the comparison networks and NeuCube with their original code versus that integrated into the custom master training files was also compared and confirmed to be comparable before the main testing began.

2.3.3 Training

In the preamble to both custom master training files - one for the comparison models and one for NeuCube - several variables can be set that allow for easy customization of the current training. In the comparison models training file, the network, dataset, training paradigm, number of epochs, and batch size can be set at the beginning of the file. The dataset, training paradigm, and the number of epochs can also be set at the beginning of the NeuCube training file. The sections of the code in these files which come from the original comparison network and NeuCube repositories are noted in the code itself.

To enable visualization of network training and performance, accuracy and loss graphing was implemented for both networks in the epoch training paradigm. Loss was calculated with the cross entropy loss function from the PyTorch library (documentation available at this link). Accuracy was calculated manually with the formula below, as used in the original training code for the comparison networks:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Number of Predictions}}$$

Loss and accuracy graphing is also available for the cross validation paradigm, though here the horizontal axis reflects the different folds instead of epochs.

The scikit learn `accuracy_score` function is additionally utilized, as employed in the original NeuCube training code, to capture average accuracy across epoch and across fold for the epoch training and cross validation training paradigms respectively.

Note that because the losses and accuracies are on significantly different scales, the produced graphs are usable as a quick reference tool for checking network training, but are not ideal for reporting purposes. The charts presented in this work were instead created separately using the .csv

files created at the end of each training round which record all losses and accuracies for the duration of training.

A confusion matrix printout was also implemented for both networks in both training paradigms, using the `confusion_matrix` function of the scikit learn package (documentation available at this link), modeling after the confusion matrix code in the original NeuCube training code.

The values needed for constructing all figures are maintained off-memory in .csv files, and the losses and accuracies are aggregated in a single report .csv file at the end of training. Two master .csv files containing all predicted labels and their corresponding true labels are constructed over the entire duration of training, with the csv files containing the predicted and true labels for each individual epoch also being maintained. This allows the confusion matrix to be constructed for a single epoch of interest or across the entire training as desired. Note that when training EEGChannelNet and the LSTM models, which train with batched data, the predictions and true labels for each epoch are generated by aggregating the predictions and true labels across the batches.

The principle metric discussed in this work is basic overall classification accuracy, though F1 score, precision, and recall are also reported, as computed with the `classification_report` function of the scikit learn package (documentation available at this link). While in cases of an unbalanced dataset, or when false positives or negatives are of particular concern, other metrics such as precision, recall, the balanced F1 metric, or the Matthews Correlation Coefficient (MCC) are advisable (Chicco & Jurman, 2020; Grootswagers et al., 2017), in this case accuracy is sufficient. A balance in the data across subject and class was confirmed in the original datasets and ensured for the quartered datasets, and in this case false positives and negatives, while not desirable, do not have any consequences beyond simply reducing the correctness of the network. The same metric is also used across all networks, so any artificial boost in reported performance would aid all the networks tested, therefore not invalidating the core comparison. The validity of comparing the networks primarily based on this metric is held up by the similar evaluations provided by the four metrics for each network (see Tables 1-4).

In the case of the epoch training paradigm, the accuracy reported is the test accuracy for the model at the epoch at which it achieved its highest validation accuracy, as calculated with the formula above. In the cross validation paradigm, the accuracy reported is the average accuracy achieved across the 5 folds, as calculated with the `accuracy_score` function of the scikit learn package (documentation available at this link). The confusion matrices in the epoch training paradigm are also generated for the model's performance on the test set on the epoch at which it achieved its highest validation accuracy, while in the cross validation paradigm, the matrices are produced across the 5 folds as was done in the original NeuCube code.

A separate Python program, `Final_Report.py`, is also available in the repository for this project and outputs all metrics and constructs and saves the relevant confusion matrices, enabling the creation of a convenient overall report file capturing information across all networks for both training paradigms and for the full and quartered dataset versions.

Due to lengthier training times with NeuCube, the maximum number of epochs was set at 50 for all networks, although the comparison networks were also tested separately over the 200 epochs used in their respective papers to evaluate how they may have improved with additional training.

3 Results

3.1 Epoch Training Paradigm

Tables 1 and 2 show the classification accuracy of each model, again as measured on the test set at the epoch at which the model achieved its highest accuracy on the validation set, in addition to their F1 score, precision, and recall. Table 1 contains these metrics for training with the full data. Metrics for training with the quartered data are contained in Table 2.

When provided with the full data and allowed to train over multiple epochs, EEGChannelNet achieved the highest accuracy score of 34.38%, with NeuCube taking third place with an accuracy score of 21.78%. The 10-layer LSTM scored the lowest with an accuracy of 0.91%. When provided with only a quarter of the data, however, NeuCube outperformed the other models with an accuracy of 16.47%, with the 5-layer LSTM taking second place with 6.85%. The 10-layer LSTM again scored the lowest with an accuracy of 1.41%. The epoch at which the reported accuracy was achieved appears in parentheses next to each score.

The second columns of both tables show the accuracies achieved when allowing the comparison models to train over 200 epochs instead of 50. While this resulted in improved accuracies for some of the networks, it did not affect NeuCube’s ranks with either the full or the quartered dataset. NeuCube’s performance is reproduced in these columns for convenient comparison.

	Accuracy (50)	Accuracy (200)	F1 Score	Precision	Recall
EEGChannelNet	34.38 (45)	40.42 (51)	0.34	0.36	0.34
LSTM	16.18 (21)	16.78 (14)	0.16	0.18	0.16
LSTM 5	28.38 (34)	28.33 (101)	0.28	0.31	0.28
LSTM 10	0.91 (5)	1.81 (1)	0	0	0.01
NeuCube	21.78 (4)	21.78 (4)	0.22	0.24	0.22

Table 1: Metrics for Epoch Training with the Full Data. Accuracy (50) and Accuracy (200) refer to the test accuracy obtained at the epoch at which the model attained its highest validation accuracy when training for 50 and 200 epochs respectively, with the epoch at which the network attained this value noted in parentheses.

	Accuracy (50)	Accuracy (200)	F1 Score	Precision	Recall
EEGChannelNet	6.45 (40)	6.05 (121)	0.07	0.09	0.06
LSTM	8.06 (33)	5.85 (34)	0.08	0.10	0.08
LSTM 5	6.85 (41)	6.45 (195)	0.07	0.08	0.07
LSTM 10	1.41 (2)	1.01 (4)	0	0	0.01
NeuCube	16.47 (7)	16.47 (7)	0.17	0.20	0.16

Table 2: Metrics for Epoch Training with the Quartered Data. Accuracy (50) and Accuracy (200) refer to the test accuracy obtained at the epoch at which the model attained its highest validation accuracy when training for 50 and 200 epochs respectively, with the epoch at which the network attained this value noted in parentheses.

Figures 3 and 4 contain the confusion matrices for all networks for the epoch training paradigm, with the full and with the quartered dataset respectively. When allowed to train on the full dataset across multiple epochs (Figure 3), all but the 10-layer LSTM show evidence of accurate predictions along the diagonal, indicating they were learning to distinguish the classes with varying degrees of success, with the pattern stronger (as expected) for those networks with higher performance based on the metrics. However, when only allowed to train with a quarter of the data (Figure 4), the confusion matrices for the comparison networks become scattered, while NeuCube’s continues to display the diagonal pattern. This indicates that the comparison networks struggled to learn the underlying patterns in the class-determined activity with the limited data, while NeuCube managed to decipher these patterns despite the reduced data quantity.

Figures 5 and 6 contain the loss and accuracy plots (plotted across epoch) for each network in the epoch training paradigm, both with the full and with the quartered dataset. With the comparison networks, loss typically fluctuated in the early epochs before beginning to settle onto a plateau beyond which they could no longer increase their performance, a pattern supported by the limited improvements seen when expanding the training duration for these networks to 200 epochs. For NeuCube, loss and accuracy fluctuate far less, with the network achieving its maximum performance early on and remaining around that figure for following epochs. Note that because NeuCube does not use batching, the range of losses for this network is substantially different from the comparison networks. The base 10 logarithm of the losses are displayed in this figure, but between

NeuCube and the comparison networks, the only appropriate comparison is the pattern of losses over time in this case, not the exact values themselves.

3.2 5-Fold Cross Validation

Tables 3 and 4 show the average classification accuracy, F1 score, precision, and recall of each model in the cross validation paradigm. Table 3 shows these metrics for training with the full data. Metrics for training for the quartered data are contained in Table 4.

With both the full and the quartered data, NeuCube outperformed the other models in this paradigm, achieving an average accuracy of 21.58% with the full dataset and 17.47% with the quartered dataset, with the other networks clustering around chance-level performance (2.5% with 40 classes with balanced data), with the exception of the single-layer LSTM, which managed a slightly above-chance accuracy of 4.70% when working with the full dataset and 3.01% when working with the quartered dataset.

	Accuracy (50)	Accuracy (200)	F1 Score	Precision	Recall
EEGChannelNet	2.52	2.77	0.01	0.01	0.03
LSTM	4.7	4.67	0.04	0.04	0.05
LSTM 5	2.21	2.48	0.01	0	0.02
LSTM 10	2.23	2.22	0	0	0.02
NeuCube	21.58	21.58	0.22	0.22	0.22

Table 3: Metrics for 5K Cross Validation Training with the Full Data. Accuracy values represent the average accuracy obtained across the 5 folds on the training rounds in which the epoch count was set to 50 and when it was set to 200, though note that in this case the epoch value is not relevant as the networks trained in a one-shot paradigm on each fold.

	Accuracy (50)	Accuracy (200)	F1 Score	Precision	Recall
EEGChannelNet	2.4	1.87	0.01	0.01	0.02
LSTM	3.01	3.07	0.03	0.03	0.03
LSTM 5	2.13	1.79	0.01	0.01	0.02
LSTM 10	2.06	2.36	0	0	0.02
NeuCube	17.47	17.47	0.18	0.18	0.17

Table 4: Metrics for 5K Cross Validation Training with the Quartered Data. Accuracy values represent the average accuracy obtained across the 5 folds on the training rounds in which the epoch count was set to 50 and when it was set to 200, though note that in this case the epoch value is not relevant as the networks trained in a one-shot paradigm on each fold.

Figures 7 and 8 contain the confusion matrices for all networks for the cross validation paradigm, both with the full and with the quartered dataset. When training under the cross validation paradigm, and therefore in essence only being allowed to train for one epoch, the comparison networks fail to show the diagonal pattern in their confusion matrices that would indicate they had learned the underlying pattern, with their matrices in this case indicating they simply continually guessed a limited number of classes. NeuCube’s confusion matrix in this condition shows the diagonal pattern, indicating that it managed to decipher the pattern despite the reduced training time.

Figures 9 and 10 contain the loss and accuracy plots for the networks across the 5 folds, both with the full and with the quartered dataset. With a few exceptions, the comparison networks display relatively consistent results across each fold, indicating that they achieved largely the same initial performance with each initialization. NeuCube displays the same pattern, but achieves a higher accuracy than the other networks in each fold. Note that once again, the base 10 logarithm of the losses is used here, and only the pattern of loss over fold may be compared between NeuCube and the comparison networks due to their different loss ranges.

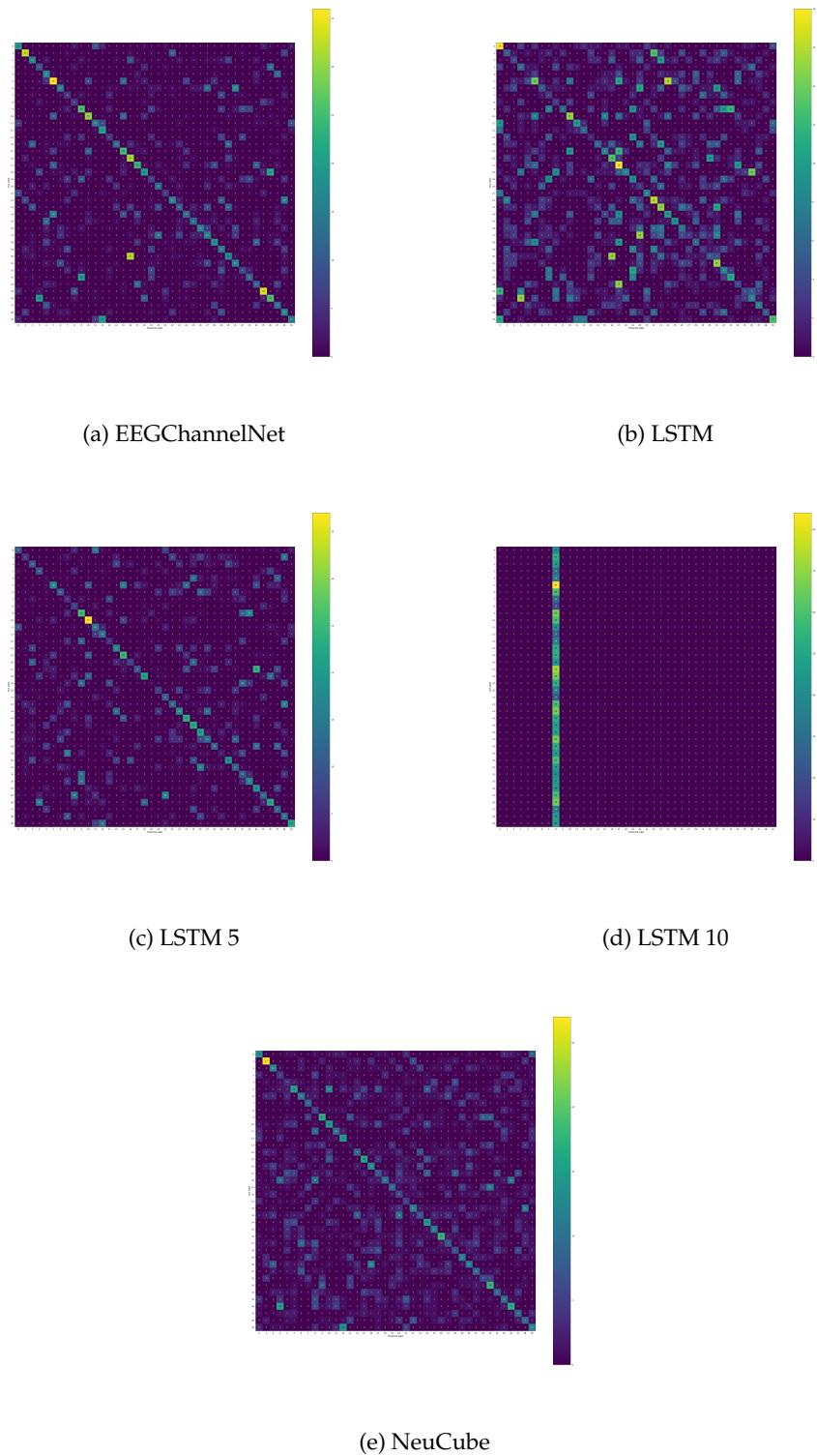


Figure 3: Confusion matrices for training with the full data in the epoch learning paradigm.

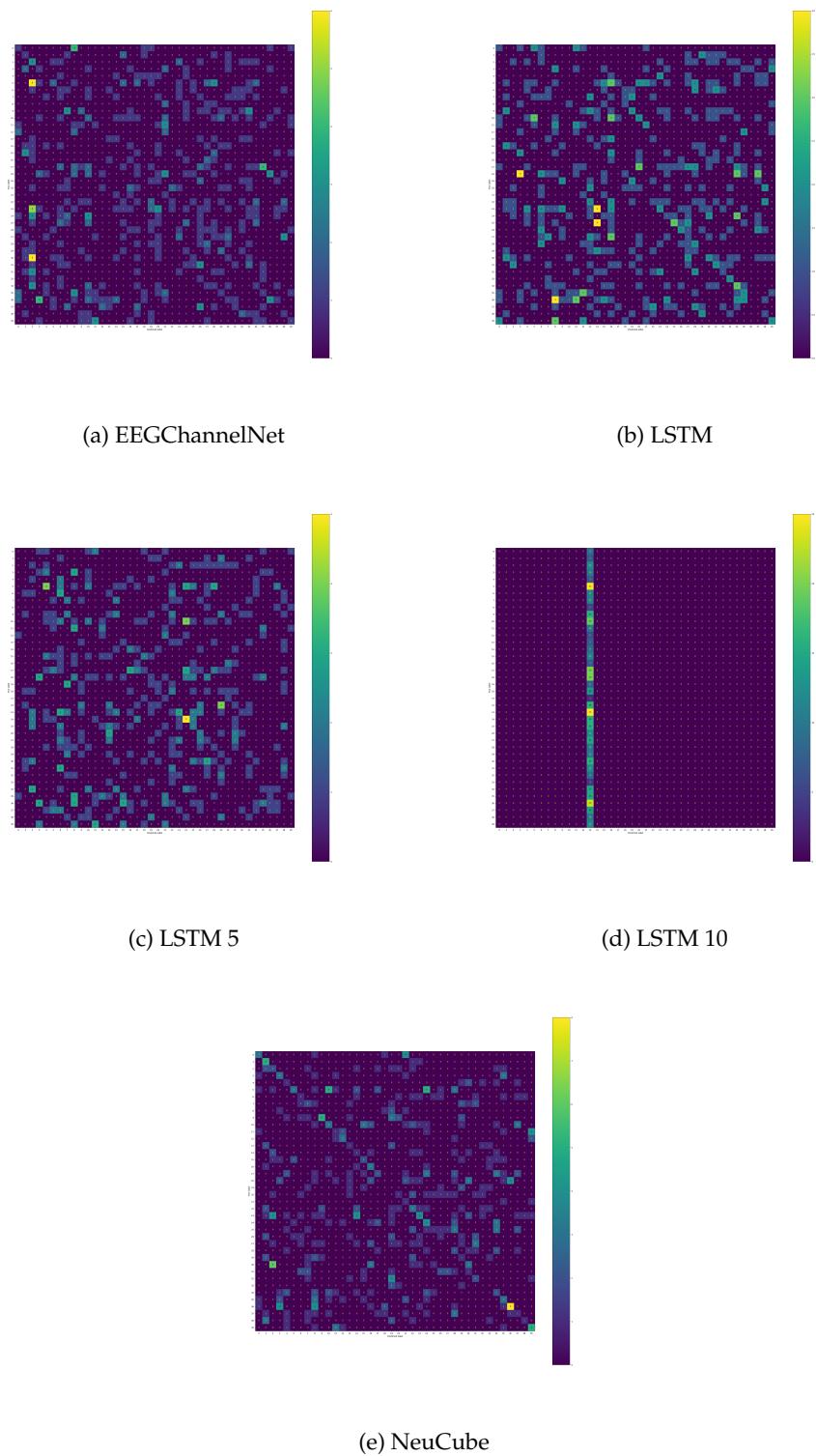


Figure 4: Confusion matrices for training with the quartered data in the epoch learning paradigm.

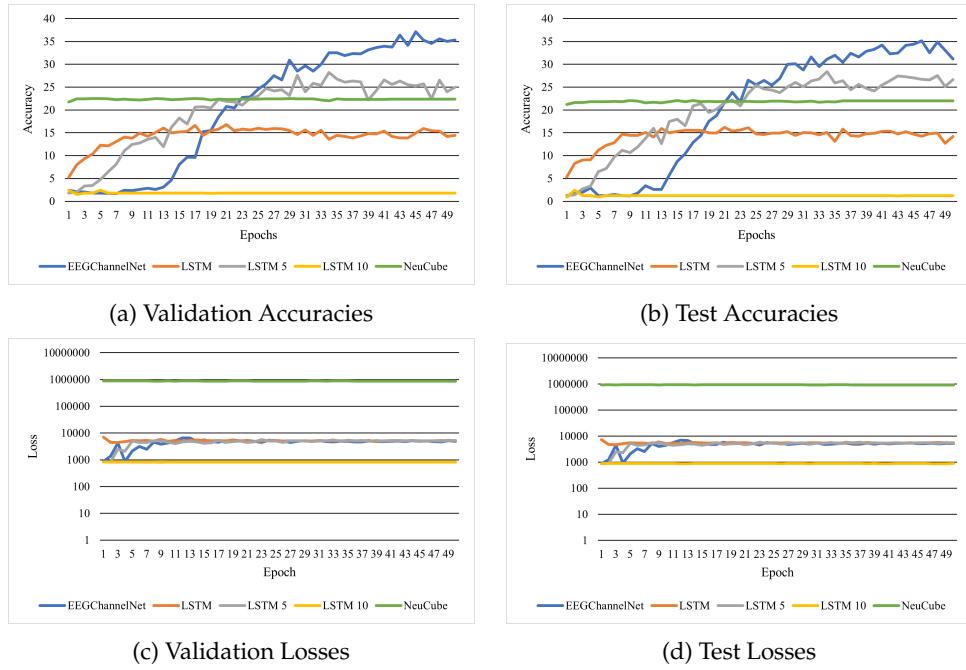


Figure 5: Loss and accuracy plots for the full data in the epoch learning paradigm.

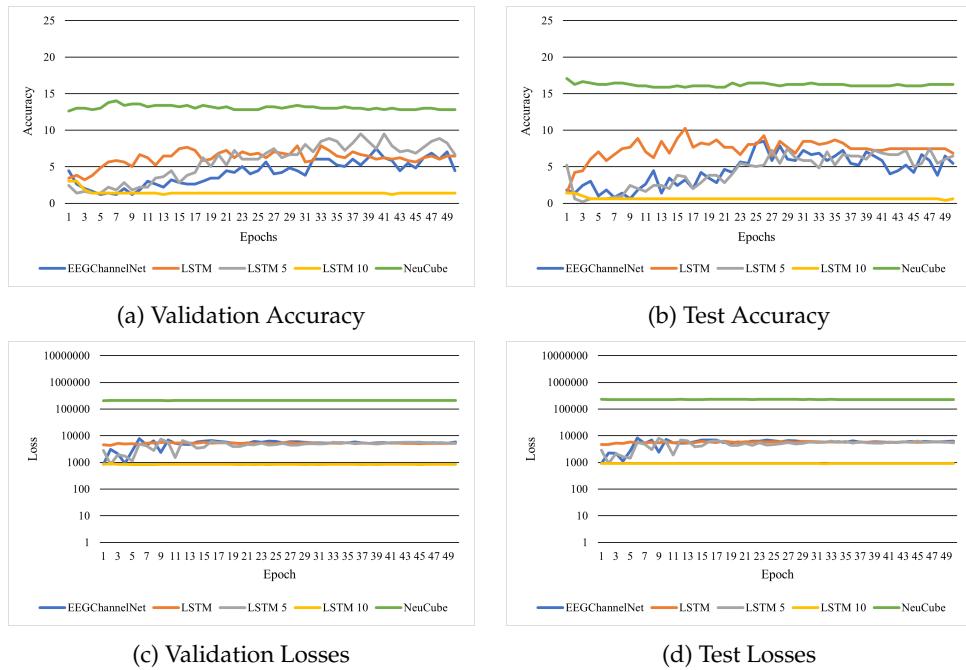


Figure 6: Loss and accuracy plots for the quartered data in the epoch learning paradigm.

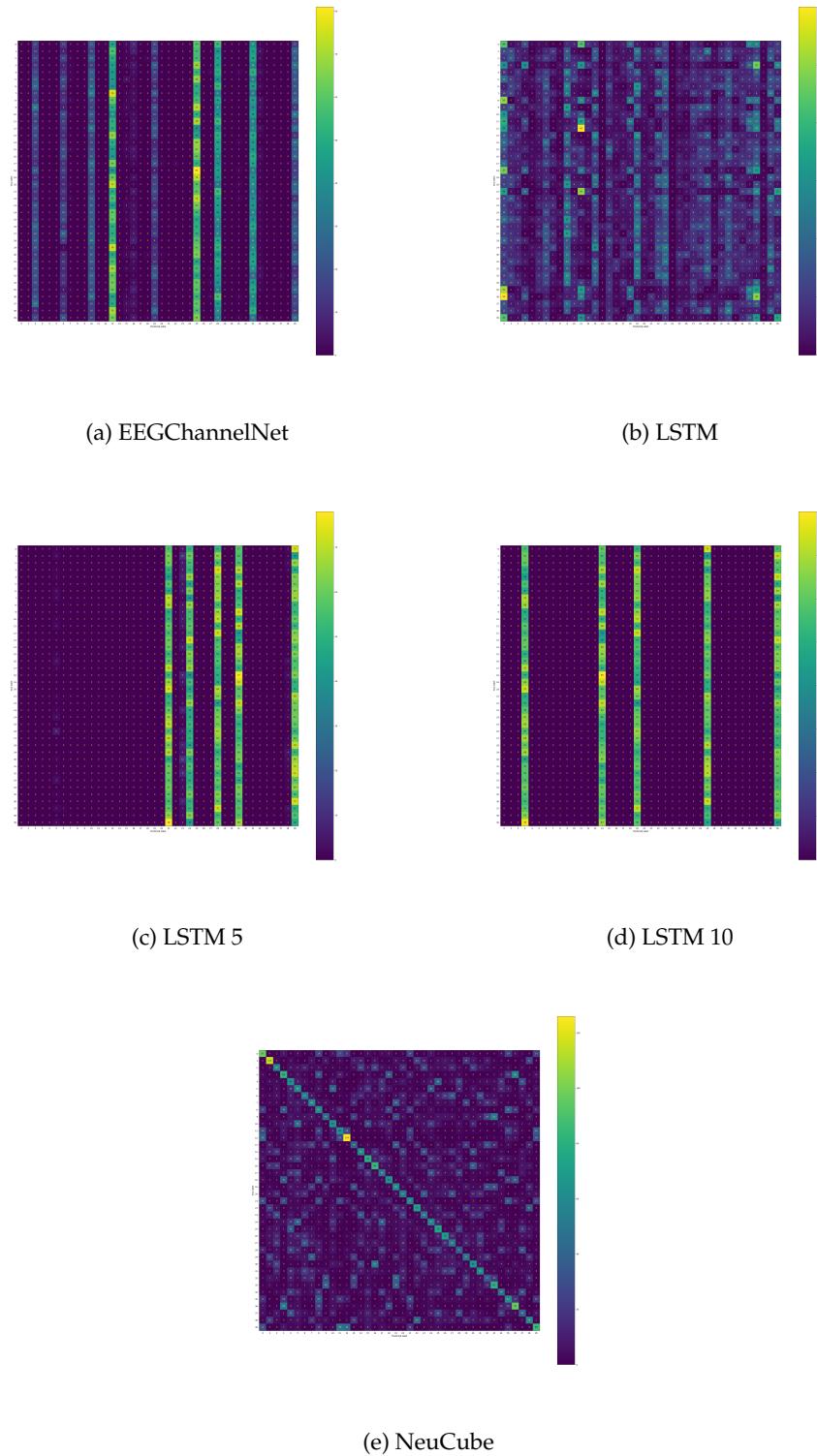


Figure 7: Confusion matrices for the cross validation paradigm with the full data.

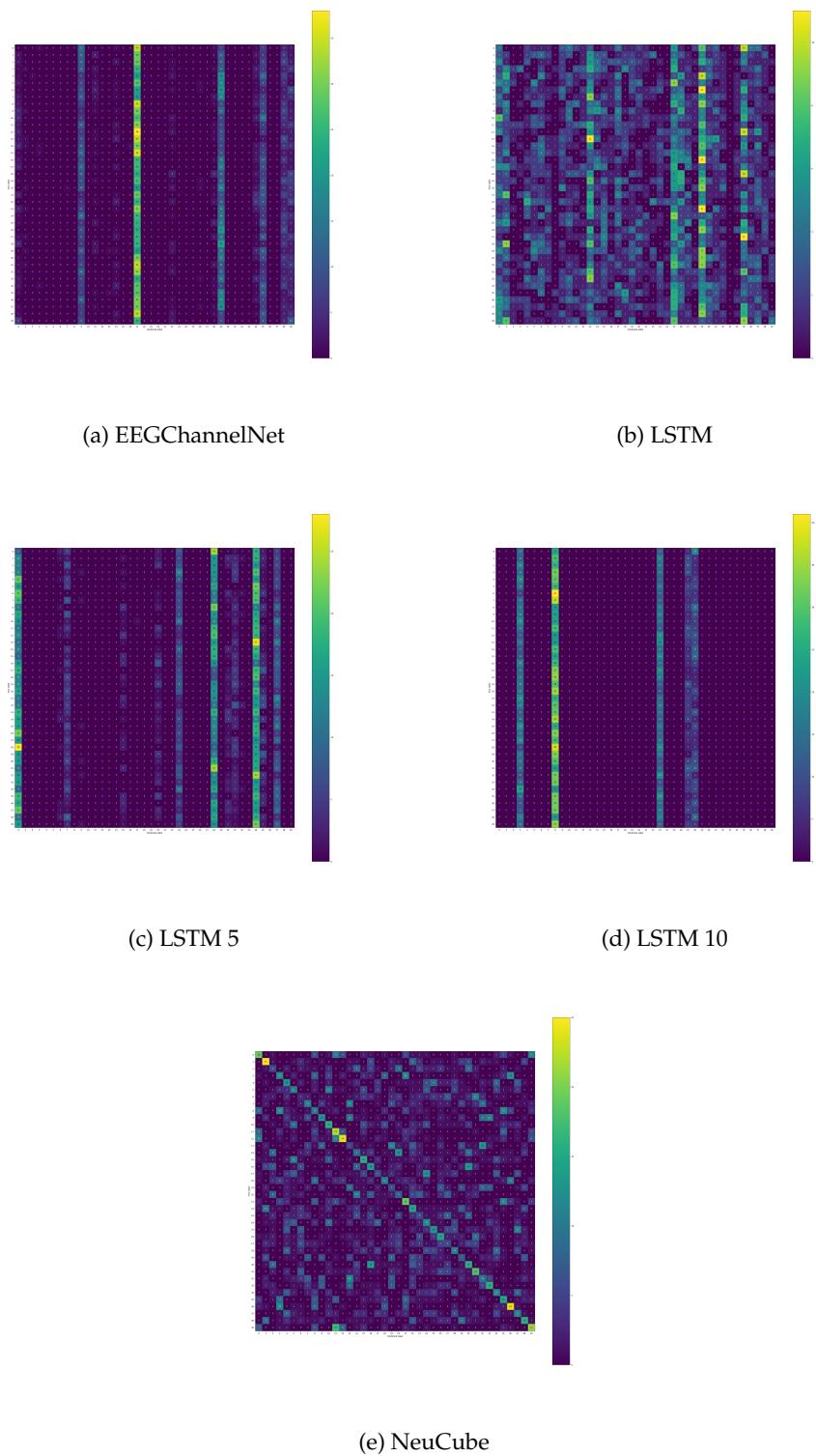


Figure 8: Confusion matrices for the cross validation paradigm with the quartered data.

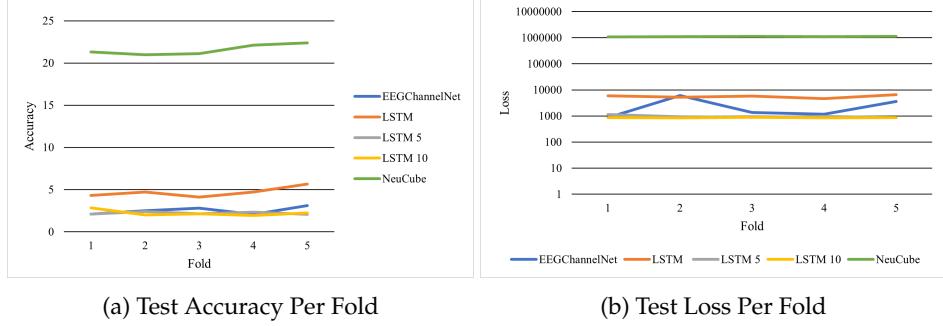


Figure 9: Loss and accuracy graphs for the cross validation paradigm with the full data.

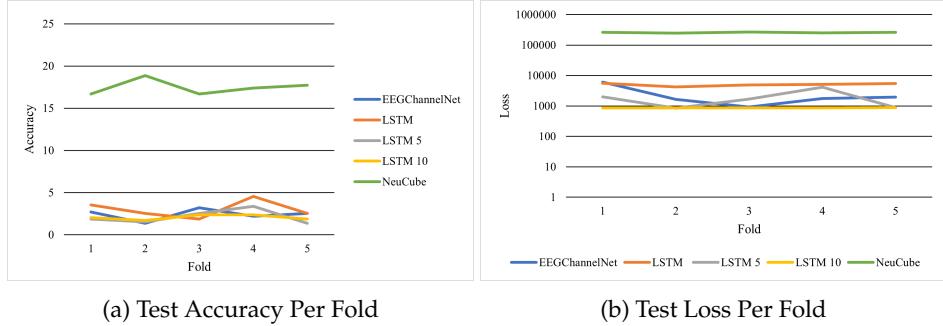


Figure 10: Loss and accuracy graphs for the cross validation paradigm with the quartered data.

3.3 Training Time

Tables 5 and 6 show the average per-epoch training time of each network. As shown, NeuCube’s per-epoch training time was slower than with the other models, with each epoch requiring roughly 40 minutes for the full data and 10 minutes for the quartered data, when trained on an NVIDIA H100 GPU system allocated 75GB of memory and 10 CPUs. However, as shown in Tables 1 and 3, NeuCube reaches its maximum accuracy faster than the other models, achieving its best accuracy within 10 epochs in the training paradigm that is never more than 1% higher than its average accuracy achieved in the 5-fold cross validation paradigm. Although due to their sub-minute per epoch training speeds the LSTM variants reach their top accuracies faster than NeuCube in terms of overall time, these accuracies are in all but the ideal case (epoch training with the full data) much lower than NeuCube’s. EEGChannelNet, the network with the best performance in the ideal condition, takes 3 hours to reach this accuracy, while NeuCube achieves close to its top performance in under one hour.

	EEGChannelNet	LSTM	LSTM 5	LSTM 10	NeuCube
Per-Epoch Training Time (Minutes)	4	1	< 1	1	40

Table 5: Training Time Full Data

	EEGChannelNet	LSTM	LSTM 5	LSTM 10	NeuCube
Per-Epoch Training Time (Minutes)	1	<1	< 1	<1	10

Table 6: Training Time Quartered Data

4 Discussion

4.1 The Validity of SNNs in Visual Classification

The general hypothesis that NeuCube, standing in here for spiking neural networks in general, would outperform the comparison networks was partially supported.

The first hypothesis, that NeuCube would universally perform better than the comparison models on the full dataset, was not supported, with EEGChannelNet being the frontrunner when allowed to train for 50 epochs on the full dataset. These findings do not support that an SNN is the best network for this task under all circumstances, particularly when a training situation may be considered ideal, with ample data and ample training time available. Even in this paradigm, however, NeuCube demonstrated a decent performance, with a classification accuracy only 6.34% lower than EEGChannelNet's, and it outperformed the comparison networks under less ideal conditions. When training in a one-shot paradigm or when working with the quartered data, the comparison networks see accuracy drops of over 10%, while NeuCube sees an average drop of only 4.71% when working with the quartered data and less than 1% when allowed less training. These findings indicate that SNNs are in fact valid competitors for decoding this kind of data, and in fact have an advantage when there is less data available or when one-shot training is preferred.

The utility of SNNs in this task is also supported with respect to overall training time. Although NeuCube takes longer to train per epoch, it makes up for this by achieving its best accuracy early on in training, and by achieving a similar accuracy on just one pass through the data. The comparison models were faster per epoch, but took multiple epochs to achieve their best results. These findings suggest SNNs may offer serious advantages not only in terms of overall performance, but also in terms of their speed in reaching that performance, an aspect which could make them attractive for a variety of use cases. An obvious application would be an EEG-based BCI, in which an SNN model as the backbone might allow adaptation to new users in less time than other models (or at least than those similar to the ones discussed here).

The point of resources becomes even more interesting when considering that in this case, the SNN was not training on the ideal system for this kind of network. In certain circumstances, SNNs can train significantly faster on neuromorphic hardware, which operates on a similar spiking principle to the SNNs themselves (Roy et al., 2019), and they are also many times more energy efficient on this class of hardware. A network deployed on Intel's Loihi 2 neuromorphic chip, for example, might train at an energy cost of less than 1 watt, significantly less than what is required to turn on a standard household light bulb, whereas even the best GPU systems can easily consume tens or even hundreds of watts (Davies et al., 2021). These points make SNNs even more attractive to BCI applications, since an SNN deployed on a chip like Loihi, or on one of any number of other neuromorphic systems, would potentially offer the same or better level of performance as a non-neuromorphic solution, and at a greatly reduced resource cost.

In comparison with the near-perfect results reported in some fMRI decoding work (e.g. Kay et al., 2008) and binary classification EEG work (e.g. Hramov et al., 2017), the accuracies themselves may feel lackluster. The maximum accuracy achieved is after all less than 50%, though this is still well above chance for this dataset. This level of accuracy is, however, not inconsistent with other work with multi-class visual classification with EEG (e.g. Kalafatovich et al., 2020; Lee et al., 2020), and in this case, may be partially the result of flaws in the dataset itself.

4.2 Issues with the Dataset

The dataset used in this study, while one of relatively few that are publicly available with thorough documentation and provided code for comparison methods, has been the subject of methodological controversy. A first point of concern raised originally in this work are the images shown to the participants, with additional methodological questions raised by a critique paper by Li et al., 2020 in 2020.

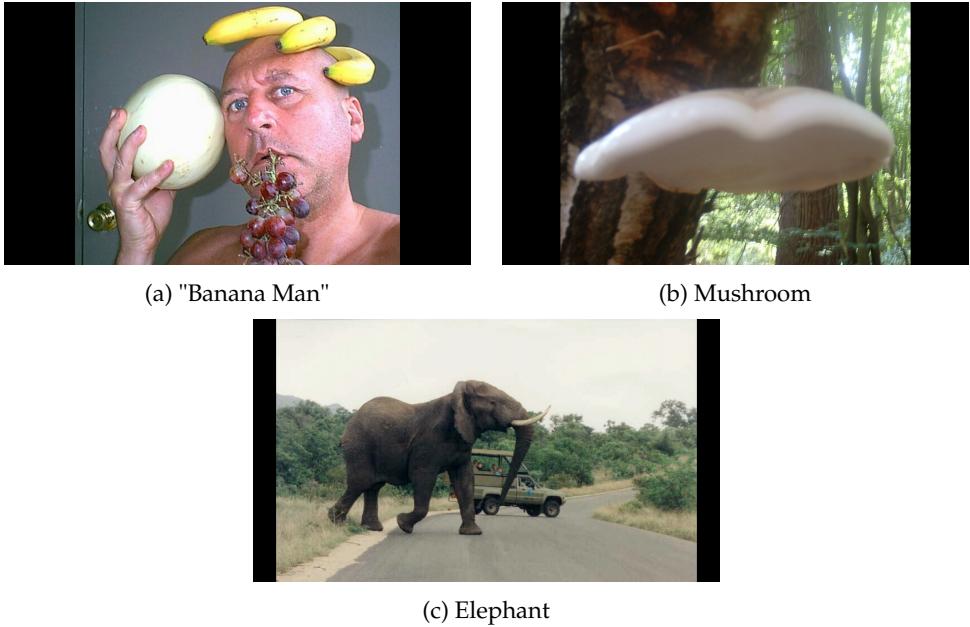


Figure 11: Example images from the dataset used in this work.

4.2.1 The Banana Man Problem

In their 2021 publication, Palazzo et al. state that the images selected for use in the dataset were chosen based on the following criteria, here directly quoted from the original paper (Palazzo et al., 2021):

- The object classes should be known and recognizable by all the subjects with a single glance;
- The object classes should be conceptually distinct and distant from each other (e.g., dog and cat categories are a good choice, whereas German shepherd and Dalmatian are not);
- The images corresponding to an object class should occupy a large portion of the image, and background of the image should be minimally complex (e.g., no other salient or distracting objects in the image). (p. 3840)

An informal review of the images in the dataset, as made available by the critique paper Li et al., 2020, reveals that this was not the case. While not all images are problematic, some have problematically poor quality (see fig. 11b), or in the case of "Banana Man" (fig. 11a), are entirely and comically inappropriate for this use case. In the former image, a person under the best circumstances might struggle to identify precisely what the image is supposed to be of in just 0.5 seconds, and the latter case is even worse.

Prior research has amply demonstrated that the brain responds uniquely to faces (Ramanathan et al., 2010) and to unexpected stimuli (Khouri & Nelken, 2015). Given this, regardless of one's opinions on the usefulness of bananas as hats, the banana itself is liable not to be seen at all in the 0.5 seconds participants had to view the image in 11a. It is therefore highly doubtful that participants viewing this image were generating EEG activity related to the actual class being tested, in this case "banana."

Unfortunately, these are only two of a number of problematic examples found in the dataset, which raises particular scientific concerns as the true nature of the images used is a clear deviation from what is stated in the papers presenting work with this dataset (e.g. Palazzo et al., 2021; Spampinato et al., 2017).

4.2.2 Methodological Concerns

Additional concerns about the experimental paradigm used to create the dataset were detailed by Li et al., 2020. The crux of their criticism is that the block design employed by the authors of the original dataset meant all stimuli for each class were presented together, which could introduce background temporal similarities to all samples of each class that would assist a classifier in assigning samples from the same category in the same dataset to that category, basing its assignment not on true class-dependent neural activity but on temporal artifacts. They subsequently performed a study of their own in which stimuli from different classes were mixed (Li et al., 2020).

The authors subsequently defended their dataset in Palazzo et al., 2020, running additional tests which determined that the temporal correlations present were minor. They further criticized the critique paper's own study by pointing out that Li et al., 2020's design resulted in participant's being on task for 23 minutes, much longer than recommended, and also reported an inability to replicate the findings of the critique paper as related to temporal contamination except when deliberately introducing a temporal correlation. They also point out that the critiquing team's decision to analyze data within-subject versus across-subject, which is indeed non-standard, also risks artificial boosting of accuracy as a result of the testing and training sets coming from the same subject in the same experimental session, and therefore both exhibiting any background temporal correlations or artifacts present during that session.

While the counterarguments of the original dataset team are convincing, and the work of the critiquing team is also not above methodological criticism, the matter of the images themselves cannot be dismissed. For the purposes of this work, it was ultimately decided that while the dataset certainly raises enough concerns to make it a questionable choice for researching visual data classification in general, it nonetheless provides a good foundation for comparing the performance of a spiking neural network (NeuCube) to other models on visual EEG data in general, as the authors openly provide both the data and the architecture and training code for their work, and all networks training on the dataset would face the same conditions.

That the dataset was suitable here for the explicit purpose of validating SNNs on visual classification data does not mean that its flaws are irrelevant. Within the goals of this paper, it provided a good platform, but any work with an aim of better classifying visual data and particularly of better understanding visual classification in the brain would be advised to consider another dataset, such as the dataset of EEG responses to selected categories from the THINGS image database (Gifford et al., 2022). The authors of this dataset better selected the images used, and the THINGS dataset they pulled from is notably adapted for research in cognitive neuroscience and was curated with human studies in mind, unlike ImageNet, from which the dataset used in this study was curated. With collections like ImageNet, as convenient as they may seem, it is crucial to recall that images in these datasets were intended for computer models and methods to work with, not for the human brain, and did not undergo screening relevant to the latter purpose. While it is not the case that all images in such challenges are unsuitable for this purpose, many of them will be. The image in 11c, for example, may be fine for the category of "elephant" due to our natural tendency to focus on living beings in images (Ramanathan et al., 2010), but a picture of a man with a banana on his head and fruit hanging out of his mouth will not likely be so successful for "banana."

4.3 Insight Into Visual Classification

Although the dataset may not be ideal, the success of EEGChannelNet and NeuCube offers insights into how to best approach EEG data from visual tasks. Both EEGChannelNet and NeuCube process the data not only temporally, but also spatially, the former in the form of the temporal and spatial convolutional blocks and the latter in the form of its spiking structure and spatially matched input structure. Past research from fMRI has strongly indicated the importance of spatial activity in object recognition and visual classification (e.g. Carlson et al., 2003; Kanwisher et al., 1997; Yacoub et al., 2008) and the importance of temporal information has also been indicated by separate work (e.g. VanRullen, 2003; VanRullen et al., 2005). Additionally, EEGChannelNet explicitly includes recurrent connections, an element identified in multiple studies as potentially critical to the function of the visual system and its recognition of objects (Kietzmann et al., 2019; Spoerer et al., 2020). While not in so explicit a form as EEGChannelNet, NeuCube's reservoir also allows for recurrent connections,

since all neurons are permitted to interact with any neuron they are connected to, opening the possibility that some neurons representing higher visual structures may feed back into those representing lower ones. The success of these two networks may indicate that the best classification performance achievable with EEG on visual tasks will necessarily need to take all three elements into account - spatial, temporal, and recurrent.

The success of the single- and 5-layer LSTM models, when allowed to train on the full data for multiple epochs, also supports the importance of recurrence and memory in this task, though the failure of the 10-layer LSTM in all cases indicates that more complex recurrent networks may not be ideal for this use, perhaps being too complex to effectively fit the data and/or encountering other common issues with deeper networks, including catastrophic forgetting and the vanishing/exploding gradients problem (Alzubaidi et al., 2021). A 5-layer model appears to be the sweet spot here, achieving the maximum accuracy in the LSTM group both during the 50 and during the 200 epoch run. However, when the data is quartered, the single-layer LSTM model performs the best from this group in almost all cases, possibly because the same issues which affect the 10-layer LSTM with the full data and multiple epochs to train begin to affect the 5-layer version as well with limited training time and limited data.

4.4 Recommendations and Future Directions

The successes and limitations of this work point to multiple avenues for future work. The proposed tri-aspect method discussed above, for example, is a key avenue for further exploration, but multiple other avenues of future research are also indicated, particularly as regards the future of SNNs.

Further analyzing exactly why NeuCube performed so well in this case could be the subject of an entire additional work, and one which is highly valuable to pursue. The success of NeuCube in this case is something which, to those who have worked with SNNs, may come as a surprise. Though it is difficult to glean from published works, which as in any field are biased towards successful results, SNNs are in fact informally well known for not working as expected, or for not working at all. Determining why this one performed well would represent a beneficial contribution to the field. It is conceivable that NeuCube's success is based largely on its unique spatio-temporal input structure, which allows it to take advantage of spatial relations that are not always modelled in other SNN architectures (N. Kasabov, 2012; N. K. Kasabov, 2014), but this warrants further investigation.

NeuCube also has utility in this field not only as a classifier, but as a model. In their study published in 2015, Capecci, Kasabov, and Wang used it to model the effects of opiate addiction and treatment (Capecci et al., 2015), and previous work with the network with fMRI has also been used to explore language comprehension (N. K. Kasabov et al., 2016). Doing the same for visual data may, as with many previous modeling studies, lead to further insights into how our brains work.

4.5 Implications for Visual Imagery

Previous work already validated NeuCube and spiking systems in general on EEG data (Behrenbeck et al., 2019; N. Kasabov et al., 2013), but the success of NeuCube with the visual EEG data in this study further indicates that this class of network is suitable specifically for visual EEG data as well, with which they had not previously been tested.

Because multiple studies have highlighted key differences in neural activity between visual imagery and visual classification (Breedlove et al., 2020; Mellet et al., 2000), it is again not valid to state that the success of the SNN on visual classification here would necessarily translate to visual imagery, but these findings do indicate this as a worthy question to explore. While a visual imagery dataset was not available during the working period of this project, one has since been released that can be used as a starting point for exploring the applications of SNNs in this domain (Wilson et al., 2023).

On a more general note, while visual classification and visual imagery are not the same, both appear to rely on both spatial and temporal activity patterns. While recurrence has been less explored

in visual imagery, it is also not inconcievable that it plays an important role in this function as well. This warrants further study with visual imagery models and decoders taking advantage of combinations of these aspects, or of all three in unison. The success of EEGChannelNet and NeuCube here, while again not directly transferrable to visual imagery, indicates that such multi-dimensional networks are worth continued exploration in this regard, from within and outside the SNN family.

5 Conclusion

While NeuCube, standing in for SNNs in general, did not perform better than the comparison networks across the board, it did exhibit a higher level of performance when given less training time or less data to train on. This points to SNNs as a valid tool worthy of further development for neural decoding in Visual Classification, and worthy of exploration in Visual Imagery. Further, exploration of models utilizing combinations of temporal, spatial, and recurrent elements is recommended, as is further testing with alternative datasets.

6 Footnotes

6.1 The Role of AI-Tools in This Project

In the modern era, it is becoming increasingly relevant and responsible to report how AI-assistance tools such as Chat-GPT were, and were not, used in projects. In the case of this project, Chat-GPT itself was not used, but a similar technology, Bing Chat (also based on the GPT model) was utilized. This tool was not used in any way in the writing of the text, but was used in its intended research capacity to inform the creation of some of the custom code presented in this work (for example, the code for processing the various .pth files), for the formatting of this document, and for locating additional sources.

6.2 Availability of Data and Code

Data and result files for this project, except where hosted by the original authors, are available via its Open Science Framework entry at this link. With the exception of the NeuCube architecture files, all code and data needed to replicate the findings of this project are available via the original data link, the OSF entry for the current project, and the code repository for this project.

All code used for this project may be found in this repository, with the exception of the code for the networks tested. The code for the comparison models can be found in the original repository for those papers, as linked in Section 2.1.2. The code for NeuCube has not yet been publicly released and is therefore not openly shared here at the request of the NeuCube team.

References

- Alazrai, R., Al-Saqqaf, A., Al-Hawari, F., Alwanni, H., & Daoud, M. I. (2020). A time-frequency distribution-based approach for decoding visually imagined objects using eeg signals. *IEEE Access*, 8, 138955–138972. <https://doi.org/10.1109/ACCESS.2020.3012918>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, 8, 1–74. <https://doi.org/10.1186/s40537-021-00444-8>
- Anupama, H., Cauvery, N., & Lingaraju, G. (2014). Real-time eeg based object recognition system using brain computer interface. *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, 1046–1051. <https://doi.org/10.1109/IC3I.2014.7019589>
- Auge, D., Hille, J., Mueller, E., & Knoll, A. (2021). A survey of encoding techniques for signal processing in spiking neural networks. *Neural Processing Letters*, 53(6), 4693–4710. <https://doi.org/10.1007/s11063-021-10562-2>
- Bagchi, S., & Bathula, D. R. (2021). Adequately wide 1d cnn facilitates improved eeg based visual object recognition. *2021 29th European Signal Processing Conference (EUSIPCO)*, 1276–1280. <https://doi.org/10.23919/EUSIPCO54536.2021.9615945>
- Bagchi, S., & Bathula, D. R. (2022). Eeg-convtformer for single-trial eeg-based visual stimulus classification. *Pattern Recognition*, 129, 108757. <https://doi.org/10.1016/j.patcog.2022.108757>
- Bang, J.-S., Jeong, J.-H., & Won, D.-O. (2021). Classification of visual perception and imagery based eeg signals using convolutional neural networks. *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, 1–6. <https://doi.org/10.48550/arXiv.2005.08842>
- Barragan-Jason, G., Cauchoux, M., & Barbeau, E. (2015). The neural speed of familiar face recognition. *Neuropsychologia*, 75, 390–401. <https://doi.org/10.1016/j.neuropsychologia.2015.06.017>
- Behrenbeck, J., Tayeb, Z., Bhiri, C., Richter, C., Rhodes, O., Kasabov, N., Espinosa-Ramos, J. I., Furber, S., Cheng, G., & Conradt, J. (2019). Classification and regression of spatio-temporal signals using neucube and its realization on spinnaker neuromorphic hardware. *Journal of Neural Engineering*, 16(2), 026014. <https://doi.org/10.1088/1741-2552/aafabc>
- Belyi, R., Gaziv, G., Hoogi, A., Strappini, F., Golan, T., & Irani, M. (2019). From voxels to pixels and back: Self-supervision in natural-image reconstruction from fmri. *Advances in Neural Information Processing Systems*, 32. <https://doi.org/10.48550/arXiv.1907.02431>
- Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., & Flor, H. (1999). A spelling device for the paralysed. *Nature*, 398(6725), 297–298. <https://doi.org/10.1038/18581>
- Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fmri and meg decoding. *Journal of Neuroscience*, 37(32), 7700–7710. <https://doi.org/10.1523/JNEUROSCI.0582-17.2017>
- Breedlove, J. L., St-Yves, G., Olman, C. A., & Naselaris, T. (2020). Generative feedback explains distinct brain activity codes for seen and mental images. *Current Biology*, 30(12), 2211–2224. <https://doi.org/10.1016/j.cub.2020.04.014>
- Capecci, E., Kasabov, N., & Wang, G. Y. (2015). Analysis of connectivity in neucube spiking neural network models trained on eeg data for the understanding of functional changes in the brain: A case study on opiate dependence treatment. *Neural Networks*, 68, 62–77. <https://doi.org/10.1016/j.neunet.2015.03.009>
- Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience*, 15(5), 704–717. <https://doi.org/10.1162/jocn.2003.15.5.704>
- Chicco, D., & Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 1–13. <https://doi.org/10.1186/s12864-019-6413-7>
- Cichy, R. M., Heinze, J., & Haynes, J.-D. (2012). Imagery and perception share cortical representations of content and location. *Cerebral Cortex*, 22(2), 372–380. <https://doi.org/10.1093/cercor/bhr106>

- Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-based fusion of meg and fmri reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cerebral Cortex*, 26(8), 3563–3579. <https://doi.org/10.1093/cercor/bhw135>
- Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the time course of individual objects with meg. *Cerebral Cortex*, 25(10), 3602–3612. <https://doi.org/10.1093/cercor/bhu203>
- Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, 105, 165–176. <https://doi.org/10.1016/j.neuropsychologia.2017.02.013>
- Davies, M., Wild, A., Orchard, G., Sandamirskaya, Y., Guerra, G. A. F., Joshi, P., Plank, P., & Risbud, S. R. (2021). Advancing Neuromorphic Computing With Loihi: A Survey of Results and Outlook. *Proceedings of the IEEE*, 109(5), 911–934. <https://doi.org/10.1109/JPROC.2021.3067593>
- Dijkstra, N., Bosch, S. E., & van Gerven, M. A. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *Journal of Neuroscience*, 37(5), 1367–1373. <https://doi.org/10.1523/jneurosci.3022-16.2016>
- Dijkstra, N., & Fleming, S. (2021). Fundamental constraints on distinguishing reality from imagination. <https://doi.org/http://dx.doi.org/10.31234/osf.io/bw872>
- Doborjeh, M. G., Capecci, E., & Kasabov, N. (2014). Classification and segmentation of fmri spatio-temporal brain data with a neucube evolving spiking neural network model. *2014 IEEE Symposium on Evolving and Autonomous Learning Systems (EALS)*, 73–80. <https://doi.org/http://dx.doi.org/10.1109/EALS.2014.7009506>
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293(5539), 2470–2473. <https://doi.org/10.1126/science.1063414>
- Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R. J., & Rees, G. (2008). Fmri activity patterns in human loc carry information about object exemplars within category. *Journal of Cognitive Neuroscience*, 20(2), 356–370. <https://doi.org/10.1162/jocn.2008.20019>
- Fulford, J., Milton, F., Salas, D., Smith, A., Simler, A., Winlove, C., & Zeman, A. (2018). The neural correlates of visual imagery vividness—an fmri study and literature review. *Cortex*, 105, 26–40. <https://doi.org/10.1016/j.cortex.2017.09.014>
- Gerstner, W. (2001). What is different with spiking neurons? *Plausible Neural Networks for Biological Modelling*, 23–48. <https://lcnwww.epfl.ch/gerstner/PUBLICATIONS/DifferentWSpiking.pdf>
- Gifford, A. T., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). A large and rich eeg dataset for modeling human visual object recognition. *NeuroImage*, 264, 119754. <https://doi.org/10.1016/j.neuroimage.2022.119754>
- Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience*, 29(4), 677–697. https://doi.org/10.1162/jocn_a_01068
- Gruber, T., Maess, B., Trujillo-Barreto, N. J., & Müller, M. M. (2008). Sources of synchronized induced gamma-band responses during a simple object recognition task: A replication study in human meg. *Brain Research*, 1196, 74–84. <https://doi.org/10.1016/j.brainres.2007.12.037>
- Halme, H.-L., & Parkkonen, L. (2016). Comparing features for classification of meg responses to motor imagery. *Plos One*, 11(12), e0168766. <https://doi.org/10.1371/journal.pone.0168766>
- Harel, A., Groen, I. I., Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The temporal dynamics of scene processing: A multifaceted eeg investigation. *eNeuro*, 3(5). <https://doi.org/10.1523%2FENEURO.0139-16.2016>
- Hebart, M. N., & Hesselmann, G. (2012). What visual information is processed in the human dorsal stream? *Journal of Neuroscience*, 32(24), 8107–8109. <https://doi.org/10.1523/JNEUROSCI.1462-12.2012>
- Herzog, M. H., & Clarke, A. M. (2014). Why vision is not both hierarchical and feedforward. *Frontiers in Computational Neuroscience*, 8, 135. <https://doi.org/10.3389/fncom.2014.00135>

- Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, 340(6132), 639–642. <https://doi.org/10.1126/science.1234330>
- Hramov, A. E., Maksimenko, V. A., Pchelintseva, S. V., Runnova, A. E., Grubov, V. V., Musatov, V. Y., Zhuravlev, M. O., Koronovskii, A. A., & Pisarchik, A. N. (2017). Classifying the perceptual interpretations of a bistable image using eeg and artificial neural networks. *Frontiers in Neuroscience*, 11, 674. <https://doi.org/10.3389/fnins.2017.00674>
- Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3(7), 4–4. <https://doi.org/10.1167/3.7.4>
- Kalafatovich, J., Lee, M., & Lee, S.-W. (2020). Decoding visual recognition of objects from eeg signals based on attention-driven convolutional neural network. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2985–2990. <https://doi.org/10.1109/SMC42975.2020.9283434>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302–4311. <https://doi.org/10.1523/jneurosci.17-11-04302.1997>
- Kaplan, R. M. (2011). The Mind Reader: the Forgotten Life of Hans Berger, Discoverer of the EEG [PMID: 21443395]. *Australasian Psychiatry*, 19(2), 168–169. <https://doi.org/10.3109/10398562.2011.561495>
- Kasabov, N. (2012). Neucube evospike architecture for spatio-temporal modelling and pattern recognition of brain signals. *Artificial Neural Networks in Pattern Recognition: 5th INNS IAPR TC 3 GIRPR Workshop, ANNPR 2012, Trento, Italy, September 17–19, 2012. Proceedings*, 5, 225–243. https://doi.org/10.1007/978-3-642-33212-8_21
- Kasabov, N., Hu, J., Chen, Y., Scott, N., & Turkova, Y. (2013). Spatio-temporal eeg data classification in the neucube 3d snn environment: Methodology and examples. *Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3–7, 2013. Proceedings, Part III* 20, 63–69. https://doi.org/10.1007/978-3-642-42051-1_9
- Kasabov, N. K. (2014). Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks*, 52, 62–76. <https://doi.org/10.1016/j.neunet.2014.01.006>
- Kasabov, N. K., Dotorjeh, M. G., & Dotorjeh, Z. G. (2016). Mapping, learning, visualization, classification, and understanding of fmri data in the neucube evolving spatiotemporal data machine of spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 28(4), 887–899. <https://doi.org/10.1109/TNNLS.2016.2612890>
- Kaspar, K., Hassler, U., Martens, U., Trujillo-Barreto, N., & Gruber, T. (2010). Steady-state visually evoked potential correlates of object recognition. *Brain Research*, 1343, 112–121. <https://doi.org/10.1016/j.brainres.2010.04.072>
- Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185), 352–355. <https://doi.org/10.1038/nature06713>
- Khouri, L., & Nelken, I. (2015). Detecting the unexpected. *Current Opinion in Neurobiology*, 35, 142–147. <https://doi.org/10.1016/j.conb.2015.08.003>
- Kietzmann, T. C., Spoerer, C. J., Sørensen, L. K., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, 116(43), 21854–21863. <https://doi.org/10.1073/pnas.1903.05946>
- Kietzmann, T. C., Ehinger, B., Porada, D., Engel, A., & König, P. (2013). From stimulus onset to category selectivity in 100ms: Category-selective visually evoked responses as a result of extensive category learning. *Journal of Vision*, 13(9), 662–662. <https://doi.org/10.1167/13.9.662>
- Knauff, M., Kassubek, J., Mulack, T., & Greenlee, M. W. (2000). Cortical activation evoked by visual mental imagery as measured by fmri. *NeuroReport*, 11(18), 3957–3962. <https://doi.org/10.1097/00001756-200012180-00011>
- Koessler, L., Maillard, L., Benhadid, A., Vignal, J. P., Felblinger, J., Vespiagnani, H., & Braun, M. (2009). Automated cortical projection of eeg sensors: Anatomical correlation via the international 10–10 system. *NeuroImage*, 46(1), 64–72. <https://doi.org/10.1016/j.neuroimage.2009.02.006>

- Kosmyna, N., Lindgren, J. T., & Lécuyer, A. (2018). Attending to visual stimuli versus performing visual imagery as a control strategy for eeg-based brain-computer interfaces. *Scientific Reports*, 8(1), 13222. <https://doi.org/10.1038/s41598-018-31472-9>
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Ganis, G., Sukel, K., & Alpert, N. (1999). The role of area 17 in visual imagery: Convergent evidence from pet and rtms. *Science*, 284(5411), 167–170. <https://doi.org/10.1126/science.284.5411.167>
- Kreiman, G., Koch, C., & Fried, I. (2000). Imagery neurons in the human brain. *Nature*, 408(6810), 357–361. <https://doi.org/10.1038/35042575>
- Kurkin, S., Chholak, P., Niso, G., Frolov, N., & Pisarchik, A. (2020). Using artificial neural networks for classification of kinesthetic and visual imaginary movements by meg data. *Saratov Fall Meeting 2019: Computations and Data Analysis: from Nanoscale Tools to Brain Functions*, 11459, 18–23. <https://doi.org/10.1111/12.2563813>
- Lee, S.-H., Lee, M., Jeong, J.-H., & Lee, S.-W. (2019). Towards an eeg-based intuitive bci communication system using imagined speech and visual imagery. *2019 IEEE international conference on systems, man and cybernetics (SMC)*, 4409–4414. <https://doi.org/10.1109/SMC.2019.8914645>
- Lee, S.-H., Lee, M., & Lee, S.-W. (2020). Neural decoding of imagined speech and visual imagery as intuitive paradigms for bci communication. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(12), 2647–2659. <https://doi.org/10.1109/TNSRE.2020.3040289>
- Li, R., Johansen, J. S., Ahmed, H., Ilyevsky, T. V., Wilbur, R. B., Bharadwaj, H. M., & Siskind, J. M. (2020). The perils and pitfalls of block design for eeg classification experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1), 316–333. <https://doi.org/10.1109/TPAMI.2020.2973153>
- Luckhurst, R. (2002, January). *The Invention of Telepathy 1870-1900*. Oxford University Press. <https://global.oup.com/academic/product/the-invention-of-telepathy-9780199249626?cc=de&lang=en&>
- Luo, Y., Fu, Q., Xie, J., Qin, Y., Wu, G., Liu, J., Jiang, F., Cao, Y., & Ding, X. (2020). Eeg-based emotion classification using spiking neural networks. *IEEE Access*, 8, 46007–46016. <https://doi.org/10.1109/ACCESS.2020.2978163>
- Mellet, E., Tzourio-Mazoyer, N., Bricogne, S., Mazoyer, B., Kosslyn, S., & Denis, M. (2000). Functional anatomy of high-resolution visual mental imagery. *Journal of Cognitive Neuroscience*, 12(1), 98–109. <https://doi.org/10.1162/08989290051137620>
- Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. (1998). Reopening the mental imagery debate: Lessons from functional anatomy. *NeuroImage*, 8(2), 129–139. <https://doi.org/10.1006/nimg.1998.0355>
- Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fmri. *Neuroimage*, 53(1), 103–118. <https://doi.org/10.1016/j.neuroimage.2010.05.051>
- Mishra, R., Sharma, K., Jha, R., & Bhavsar, A. (2023). Neurogan: Image reconstruction from eeg signals via an attention-based gan. *Neural Computing and Applications*, 35(12), 9181–9192. <https://doi.org/10.1007/s00521-022-08178-1>
- Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.-a., Morito, Y., Tanabe, H. C., Sadato, N., & Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5), 915–929. <https://doi.org/10.1016/j.neuron.2008.11.004>
- Mozafari, M., Reddy, L., & VanRullen, R. (2020). Reconstructing natural scenes from fmri patterns using bigbigan. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8. <https://doi.org/10.48550/arXiv.2001.11761>
- Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage*, 105, 215–228. <https://doi.org/10.1016/j.neuroimage.2014.10.018>
- Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6), 902–915. <https://doi.org/10.1016/j.neuron.2009.09.006>

- Nishimura, K., Aoki, T., Inagawa, M., Tobinaga, Y., & Iwaki, S. (2015). Brain activities of visual thinkers and verbal thinkers: A meg study. *Neuroscience Letters*, 594, 155–160. <https://doi.org/https://doi.org/10.1016/j.neulet.2015.03.043>
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fmri data. *Trends in Cognitive Sciences*, 10(9), 424–430. <https://doi.org/https://doi.org/10.1016/j.tics.2006.07.005>
- Nowinski, W. L. (2005). The cerefy brain atlases: Continuous enhancement of the electronic talairach-tournoux brain atlas. *Neuroinformatics*, 3, 293–300. <https://doi.org/https://doi.org/10.1385/ni:3:4:293>
- Palazzo, S., Spampinato, C., Kavasidis, I., Giordano, D., Schmidt, J., & Shah, M. (2021). Decoding Brain Representations by Multimodal Learning of Neural Activity and Visual Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11), 3833–3849. <https://doi.org/10.1109/TPAMI.2020.2995909>
- Palazzo, S., Spampinato, C., Schmidt, J., Kavasidis, I., Giordano, D., & Shah, M. (2020). Correct block-design experiments mitigate temporal correlation bias in eeg classification. *bioRxiv*. <https://doi.org/https://doi.org/10.1101/2020.12.05.403402>
- Paugam-Moisy, H., & Bohte, S. M. (2012). Computing with Spiking Neuron Networks. In *Handbook of Natural Computing* (pp. 335–376). Springer Berlin Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-540-92910-9_10
- Peng, L., Hou, Z.-G., Kasabov, N., Bian, G.-B., Vladareanu, L., & Yu, H. (2015). Feasibility of neu-cube spiking neural network architecture for emg pattern recognition. *2015 International Conference on Advanced Mechatronic Systems (ICAMechS)*, 365–369. <https://doi.org/https://doi.org/10.1109/ICAMechS.2015.7287090>
- Qiao, K., Chen, J., Wang, L., Zhang, C., Zeng, L., Tong, L., & Yan, B. (2019). Category decoding of visual stimuli from human brain activity using a bidirectional recurrent neural network to simulate bidirectional information flows in human visual cortices. *Frontiers in Neuroscience*, 13, 692. <https://doi.org/https://doi.org/10.3389/fnins.2019.00692>
- Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R., & Khaligh-Razavi, S.-M. (2019). Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *PLoS Computational Biology*, 15(5), e1007001. <https://doi.org/https://doi.org/10.1371/journal.pcbi.1007001>
- Ramanathan, S., Katti, H., Sebe, N., Kankanhalli, M., & Chua, T.-S. (2010). An eye fixation database for saliency detection in images. *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part IV* 11, 30–43. https://doi.org/http://dx.doi.org/10.1007/978-3-642-15561-1_3
- Roland, P., & Gulyas, B. (1994). Visual imagery and visual representation. *Trends in Neurosciences*, 17(7), 281–287. [https://doi.org/https://doi.org/10.1016/0166-2236\(94\)90057-4](https://doi.org/https://doi.org/10.1016/0166-2236(94)90057-4)
- Roy, K., Jaiswal, A., & Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature* 575, 607–617, 575, 607–617. <https://doi.org/10.1038/s41586-019-1677-2>
- Rus, I. D., Marc, P., Dinsoreanu, M., Potolea, R., & Muresan, R. C. (2017). Classification of eeg signals in an object recognition task. *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 391–395. <https://doi.org/http://dx.doi.org/10.1109/ICCP.2017.8117036>
- Sanders, L. (2021). How Hans Berger's Quest for Telepathy Spurred Modern Brain Science [Accessed: 28.07.2023]. [https://www.sciencenews.org/article/hans-berger-telepathy-neuroscience-brain-eeg#:~:text=Chasing%20after%20a%20scientific%20basis%20for%20telepathy%20was,device%20that%20could%20read%20the%20brain%20%20%20electrical%20activity](https://www.sciencenews.org/article/hans-berger-telepathy-neuroscience-brain-eeg#:~:text=Chasing%20after%20a%20scientific%20basis%20for%20telepathy%20was,device%20that%20could%20read%20the%20brain%20%20%20%20electrical%20activity)
- Schrauwen, B., & Van Campenhout, J. (2003). Bsa, a fast and accurate spike train encoding scheme. *Proceedings of the International Joint Conference on Neural Networks*, 2003, 4, 2825–2830. <https://doi.org/https://doi.org/10.1109/IJCNN.2003.1224019>
- Shatek, S. M., Grootswagers, T., Robinson, A. K., & Carlson, T. A. (2019). Decoding images in the mind's eye: The temporal dynamics of visual imagery. *Vision*, 3(4), 53. <https://doi.org/https://doi.org/10.3390/vision3040053>
- Shen, G., Dwivedi, K., Majima, K., Horikawa, T., & Kamitani, Y. (2019). End-to-end deep image reconstruction from human brain activity. *Frontiers in Computational Neuroscience*, 13, 21. <https://doi.org/https://doi.org/10.3389/fncom.2019.00021>

- Shimizu, H., & Srinivasan, R. (2022). Improving classification and reconstruction of imagined images from eeg signals. *Plos One*, 17(9), e0274847. <https://doi.org/10.1371/journal.pone.0274847>
- Singh, P., Pandey, P., Miyapuram, K., & Raman, S. (2023). Eeg2image: Image reconstruction from eeg brain signals. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. <https://doi.org/10.1109/ICASSP49357.2023.10096587>
- Spampinato, C., Palazzo, S., Kavasidis, I., Giordano, D., Souly, N., & Shah, M. (2017). Deep Learning Human Mind for Automated Visual Classification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4503–4511. <https://doi.org/10.1109/CVPR.2017.479>
- Spoerer, C. J., Kietzmann, T. C., Mehrer, J., Charest, I., & Kriegeskorte, N. (2020). Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision. *PLoS Computational Biology*, 16(10), e1008215. <https://doi.org/10.1371/journal.pcbi.1008215>
- Stefanics, G., Jakab, A., Bernáth, L., Kellényi, L., & Hernádi, I. (2004). Eeg early evoked gamma-band synchronization reflects object recognition in visual oddball tasks. *Brain Topography*, 16, 261–264. <https://doi.org/10.1023/b:brat.0000032862.38122.b6>
- Stewart, A. X., Nuthmann, A., & Sanguinetti, G. (2014). Single-trial classification of eeg in a visual object task using ica and machine learning. *Journal of Neuroscience Methods*, 228, 1–14. <https://doi.org/10.1016/j.jneumeth.2014.02.014>
- Tafreshi, T. F., Daliri, M. R., & Ghodousi, M. (2019). Functional and effective connectivity based features of eeg signals for object recognition. *Cognitive Neurodynamics*, 13, 555–566. <https://doi.org/10.1007/s11571-019-09556-7>
- Talairach, J., Tournoux, P., & Rayport, M. (1988). Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging. *The Journal of Laryngology & Otology*, 104. <https://doi.org/10.1017/S0022215100111892>
- Tan, C., Sarlija, M., & Kasabov, N. (2020). Spiking neural networks: Background, recent development and the neucube architecture. *Neural Processing Letters*, 52, 1675–1701. <https://doi.org/10.1007/s11063-020-10322-8>
- Taylor, D., Scott, N., Kasabov, N., Capecci, E., Tu, E., Saywell, N., Chen, Y., Hu, J., & Hou, Z.-G. (2014). Feasibility of neucube snn architecture for detecting motor execution and motor intention for use in bciapplications. *2014 International Joint Conference on Neural Networks (IJCNN)*, 3221–3225. <https://doi.org/10.1109/IJCNN.2014.6889936>
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522. <https://doi.org/10.1038/381520a0>
- Tu, E., Kasabov, N., Othman, M., Li, Y., Worner, S., Yang, J., & Jia, Z. (2014). Neucube (st) for spatio-temporal data predictive modelling with a case study on ecological data. *2014 International Joint Conference on Neural Networks (IJCNN)*, 638–645. <https://doi.org/10.1109/IJCNN.2014.6889717>
- VanRullen, R. (2003). Visual saliency and spike timing in the ventral visual pathway. *Journal of Physiology-Paris*, 97(2-3), 365–377. <https://doi.org/10.1016/j.jphysparis.2003.09.010>
- VanRullen, R., Guyonneau, R., & Thorpe, S. J. (2005). Spike times make sense. *Trends in Neurosciences*, 28(1), 1–4. <https://doi.org/10.1016/j.tins.2004.10.010>
- Virgilio, G. C., Sossa, A. J., Antelis, J. M., & Falcón, L. E. (2020). Spiking neural networks applied to the classification of motor tasks in eeg signals. *Neural Networks: The Official Journal of the International Neural Network Society*, 122, 130–143. <https://doi.org/10.1016/j.neunet.2019.09.037>
- Wang, C., Hu, X., Yao, L., Xiong, S., & Zhang, J. (2011). Spatio-temporal pattern analysis of single-trial eeg signals recorded during visual object recognition. *Science China Information Sciences*, 54, 2499–2507. <https://doi.org/10.1007/s11432-011-4507-1>
- Wilson, H., Golbabaei, M., Proulx, M. J., Charles, S., & O'Neill, E. (2023). Eeg-based bci dataset of semantic concepts for imagination and perception tasks. *Scientific Data*, 10(1), 1–11. <https://doi.org/10.1038/s41597-023-02287-9>
- Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual imagery and perception share neural representations in the alpha frequency band. *Current Biology*, 30(13), 2621–2627. <https://doi.org/10.1016/j.cub.2020.04.074>

- Yacoub, E., Harel, N., & Uğurbil, K. (2008). High-field fmri unveils orientation columns in humans. *Proceedings of the National Academy of Sciences*, 105(30), 10607–10612. <https://doi.org/https://doi.org/10.1073/pnas.0804110105>
- Yavandhasani, M., & Ghaderi, F. (2021). Visual object recognition from single-trial eeg signals using machine learning wrapper techniques. *IEEE Transactions on Biomedical Engineering*, 69(7), 2176–2183. <https://doi.org/https://doi.org/10.1109/tbme.2021.3138157>
- Zeman, A. Z., Della Sala, S., Torrens, L. A., Gountouna, V.-E., McGonigle, D. J., & Logie, R. H. (2010). Loss of imagery phenomenology with intact visuo-spatial task performance: A case of 'blind imagination'. *Neuropsychologia*, 48(1), 145–155. <https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2009.08.024>
- Zhang, Y., Jia, S., Zheng, Y., Yu, Z., Tian, Y., Ma, S., Huang, T., & Liu, J. K. (2020). Reconstruction of natural visual scenes from neural spikes with deep neural networks. *Neural Networks*, 125, 19–30. <https://doi.org/https://doi.org/10.48550/arXiv.1904.13007>
- Zheng, X., & Chen, W. (2021). An attention-based bi-lstm method for visual object classification via eeg. *Biomedical Signal Processing and Control*, 63, 102174. <https://doi.org/https://doi.org/10.1016/j.bspc.2020.102174>