**UMEÅ UNIVERSITY**
**Department of Engineering Physics**
**Master's Thesis**

**November 19, 2021**

Master's Thesis in Engineering Physics

# Assessment of acute vestibular syndrome using deep learning:

## Classification based on head-eye positional data from a video head-impulse test

**Name** Hugo Johansson
**Email** hujo0004@student.umu.se

**Supervisors:** Helena Grip, Johan Skönevik
**Examiner:** Fredrik Öhberg

**Abstract**

The field of medicine is always evolving and one step in this evolution is the use of decision support systems like artificial intelligence. These systems open the possibility to minimize human error in diagnostics as practitioners can use objective measurements and analysis to assist with the diagnosis. In this study the focus has been to explore the possibility of using deep learning models to classify stroke, vestibular neuritis and control groups based on data from a video head impulse test (vHIT). This was done by pre-processing data from vHIT into features that could be used as input to an artificial neural network. Three different models were designed, where the first two used mean motion data describing the motion of the head and eyes and their standard deviations, and the last model used extracted parameters. The models were trained from vHIT-data from 76 control cases, 37 vestibular neuritis cases and 46 stroke cases. To get a better grasp of the differences between the groups, a comparison was made between the parameters and the mean curves. The resulting models performed to a varying degree with the first model correctly classified 77.8 % of the control cases, 55.6 % of the stroke cases and 80 % of the vestibular neuritis cases. The second model correctly classified 100 % of the control cases, 11.1 % of the stroke cases and 80.0 % of the vestibular neuritis cases. Lastly the third model correctly classified 77.8 % of the control cases, 22.2 % of the stroke cases and 100 % of the vestibular neuritis cases. The results are still insufficient when it comes to clinical use, as the stroke classification requires a higher sensitivity. This means that the cases are correctly classified and gets the urgent care they need. However, with more data and research, these methods could improve further and then provide a valuable service as decision support systems.

***Keywords:*** *Deep learning, vHIT, stroke, vestibular neuritis*

# Acknowledgements

I would like to begin by thanking Helena and Johan for supporting me during this project. Their help has been invaluable, and I have learnt much from our meetings and the feedback they have given both on the project and my report.

Secondly, I want to thank Fredrik Öhberg for the feedback during the writing process as it made the process much more manageable.

I also want to thank Jonatan Salzer and Anna Diamant for their help and expertise regarding vestibular neuritis and stroke.

Lastly, I want to thank my family and my friends for supporting me during this project, with special thanks to my walking buddy and the support pillars on discord.

# Contents

# 1 Introduction

In the field of medicine today technology acts as a tool for doctors and nurses to diagnose patients. This can be a magnetic resonance imaging-scan (MRI-scan) that shows an apparent stroke lesion which otherwise would have been impossible to notice. Technology continues to improve, and as consequence, new ways for it to provide a diagnosis. One of these new ways is AI decision support for diagnosis. There have been multiple studies applying AI-models to try to diagnose different conditions, such as depression detection, and vestibular neuritis (VN) diagnosis using head impulse test data. [1, 2, 3]. The VN studies handles classification between VN and control, but another diagnosis that can have similar symptoms is stroke. This gives a need for models that could differentiate between these groups.

## 1.1 Acute Vestibular syndrome – causes and current treatment and diagnosis

Vestibular syndromes can have different causes and different severities. Harmless and treatable dizziness can be caused by inflammations in the ear canal (VN) as compared to brain damage caused by inadequate blood flow to brain cells (stroke) can cause more diffuse vestibular symptoms and needs acute care to avoid severe brain damage. In a study performed by Ljunggren et. al. from 2017, reviewing records from the emergency department at the University Hospital of Umeå, showed that 2.1 % of patients had dizziness symptoms [4]. Also this study showed that around 4.8 % of these were due to cerebrovascular causes, where 2.8 % percent had specifically a stroke. Some of these cases of dizziness can be summarized under the generic term Acute Vestibular Syndromes which is characterized by acute vertigo and dizziness that lasts for more than 24 hours [5]. Two common underlying causes for this are VN and stroke, which means that there is a need for accurate methods that differentiate between these two, as they require drastically different treatments. This is usually done with clinical tests such as the Caloric reflex test, MRI-scan, and the HINTS-test (head impulse, nystagmus, and test of skew) [5]. These methods have a high sensitivity when it comes to detecting both VN and strokes with HINTS-test and MRI-scan having around 80-90 % sensitivity regarding VN and PCS(posterior circulation stroke) [5, 6]. These methods are not perfect as the MRI-scans can be costly and HINTS-test need an experienced examiner. Both MRI-scans and Caloric reflex test are basic diagnostic tests for strokes and VN [7, 8].

There is another method used today, the video head impulse test (vHIT), which is a modified version of the head impulse test (HIT). Each test works by delivering fast unpredictable head rotations in the direction of the 3 vestibular canals on each side of the head, at the same time as the patient keeps its gaze at a fixed point. This in turn activates the vestibular ocular reflex which counteract the fast head movement with an opposite movement of the eyes. This movement is tracked by an examiner in the case of the HIT, but in the vHIT case the movement is tracked by video and further analysis can be done after the patient visit. A visualization of the vestibular canals can be seen in Figure 1.

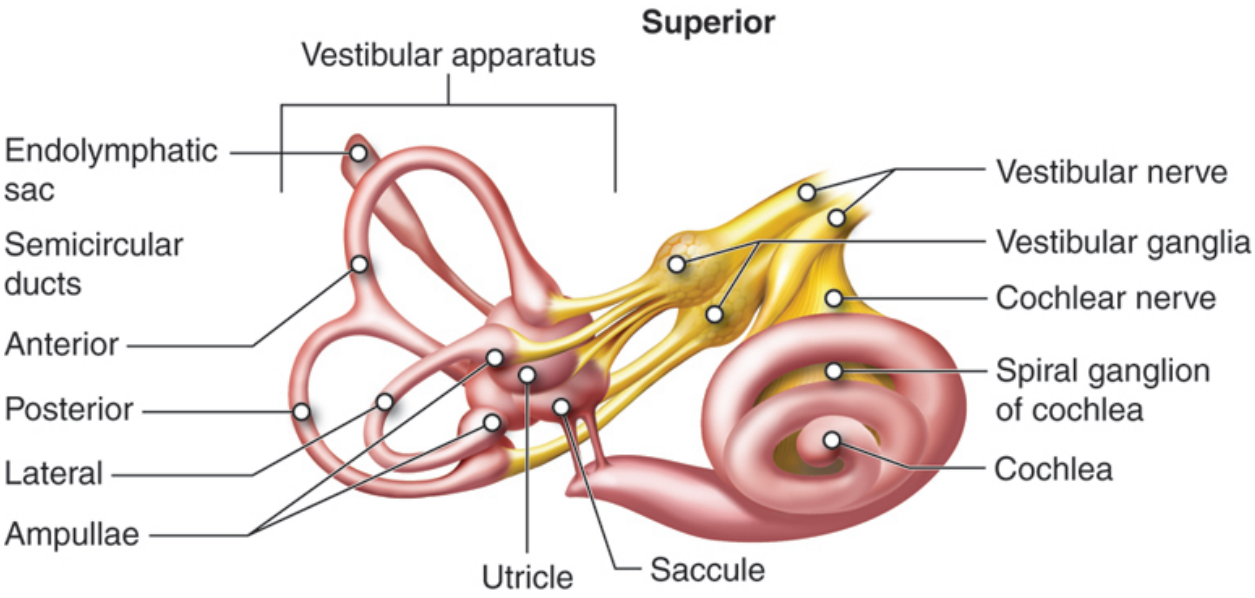



**Figure 1** – An illustration of the vestibular canals in the inner ear [9]. This work by Cenveo is licensed under a Creative Commons Attribution 3.0 United States (http://creativecommons.org/licenses/by/3.0/us/)

The biggest part of the data-analysis is calculating the gain of the different vestibular canals, which is a parameter that compares the head movement to the eye movement. These values are then indicative if the patient has normal function in their separate vestibular canals and are labelled as the vestibulo-ocular reflex gain (VOR-gain). Regarding VN and stroke, it has been shown that the gain can be very indicative if a patient has VN as shown by Guler et. al. [6]. For stroke it is more complex as their gain can be quite like a person with normal vestibular function, but by considering gain asymmetry some distinction can be made as stated in Guler et. al. To further improve the results, Guler et. al [6]. suggest including saccade analysis to capture the more complex patterns that might occur in both VN and strokes.

## 1.2 Artificial intelligence for medical applications

With the advancement of artificial intelligence methods, such as machine learning and deep learning, the search for applications has increased drastically. In the field of medicine studies have been performed to test if these methods can be used to help with data analysis and with diagnosis. This has been done from typical clinical parameters from already established tests as well as using data such as head and eye movement.

One of these studies were performed by Kacem et. al. in 2018 and it studied the use of head movements to detect if a patient suffered from depression [3]. It used encoded facial and head movement to classify severity of depression into three different levels, remitted, mildly depressed and severely depressed. It managed to reach and average accuracy of 71 % when using both facial and head movement, and even 84 % accuracy when differentiating between mild and severe depression. The study shows how using an already know fact, such as that expressiveness correlates with depression, can be applied to a machine learning models, in this case a support vector machine model, to extract information that would require further

testing.

Another study done by Landry et. al. focused on concussion and the relation between eye movement and the diagnosis [10]. They based their method around saccade parameters and chose the ones who were significantly different between the control group and the concussion and post-concussion syndrome (PCS) groups and applied these parameters to an artificial neural network (ANN) for classification. With these parameters they reached 67 % accuracy between the control group and the concussion and 72 % accuracy between the control and the PCS group. They did not manage do differentiate between the concussion group and the PCS group, which shows that similar cases can be hard to differentiate between similar cases.

Finally, Nguyen et. al. have done two studies regarding using vHIT-data to build clinical decision support systems [1, 2]. The initial system used covert catch-up saccade data and the positional gain data for the decision support application and the second study also considered artifacts present in the data. These studies showed a very practical application which tried to match a clinician's reasoning while performing the analysis.

## 1.3 Aim

As previous studies have shown there has been multiple ideas with either facial and head movement or eye movement. The aim of this study was to further analyze the ability to differentiate between VN, stroke and healthy controls based on head- and eye position data collected during a vHIT, including all six vertebral canals. This study will be done by performing data analysis on the extracted data looking at VOR-gain, saccades, and other parameters of interest. It will be further done by creating a deep learning model based on the data from the extracted data to be able to predict a diagnosis for an unknown test person. Our goal was to try out three different AI models, a model based on time series data from all six canals, one based on time series data from lateral canals and finally one based on parameters describing the head and eye movement. The hypothesis was that the models could reach a specificity and sensitivity of 80 %, similar level to the results in the studies performed by Calic and Guler [5, 6].

# 2 Theory

## 2.1 Artificial Neural Networks

### 2.1.1 Architecture

An artificial neural network (ANN) is a way to emulate the power of the brain using a computer. This is done using artificial neurons which act similarly fashion as our own neurons by taking input and transforming it to a specific output. These neurons are built up in layers where there is an input and output layer, and then hidden layers in between these, as illustrated in an example in Figure 2.

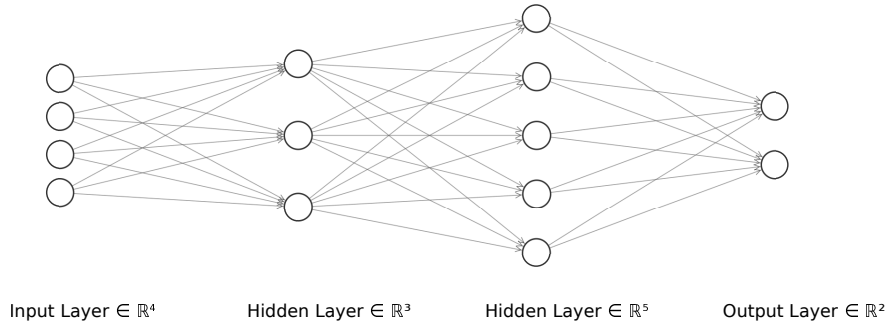



**Figure 2** – An example of an architecture of an artificial neural network with 2 hidden layers.

What connects the different layers are the numerical weights that indicate the importance of each node and they in turn work in combination with a possible bias as the input for the activation function of a node. These activation functions will be the final output of the node which can then be combined with weights for the next layer. This calculation is visualized in Figure 3.

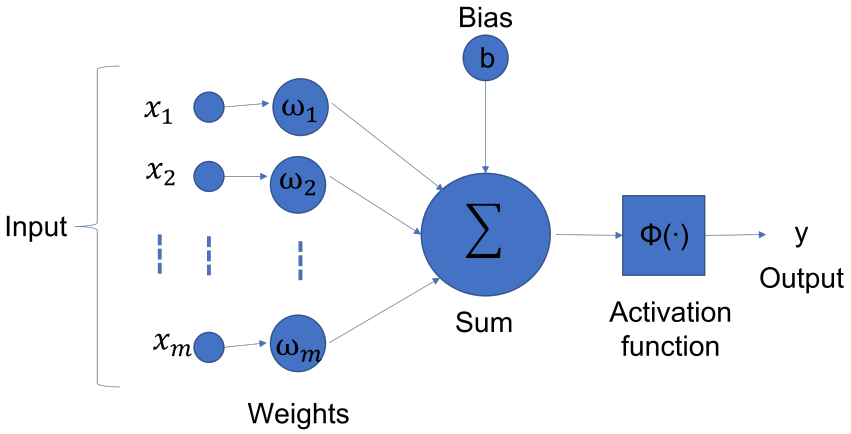



**Figure 3** – An example of the operations performed by each node in each layer, where $x_i$ is the input from each node, $\omega_i$ is the weights assigned to each node, b is the possible bias and $\Phi$ is the activation function which in turn gives the output y.

The activation functions will act as the nonlinear part of the network giving it the capability to handle nonlinear problems.

### 2.1.2 Activation functions

Throughout the evolution of neural networks, different types of activation functions have been used. The three most popular ones have been the Sigmoid function, the tanh function and the ReLU function [11]. The Sigmoid function is defined as

$$f(x) = \frac{1}{1 + e^{-x}},$$

and was widely used in the early 1990's as an activation function that maps input in the range 0 to 1 and didn't get replaced until the hyperbolic tangent function came into popularity. The tanh function turned out to make models easier to train and gave better predictive performances and it maps values within the range of -1 to 1. These functions are shown in Figure 4.

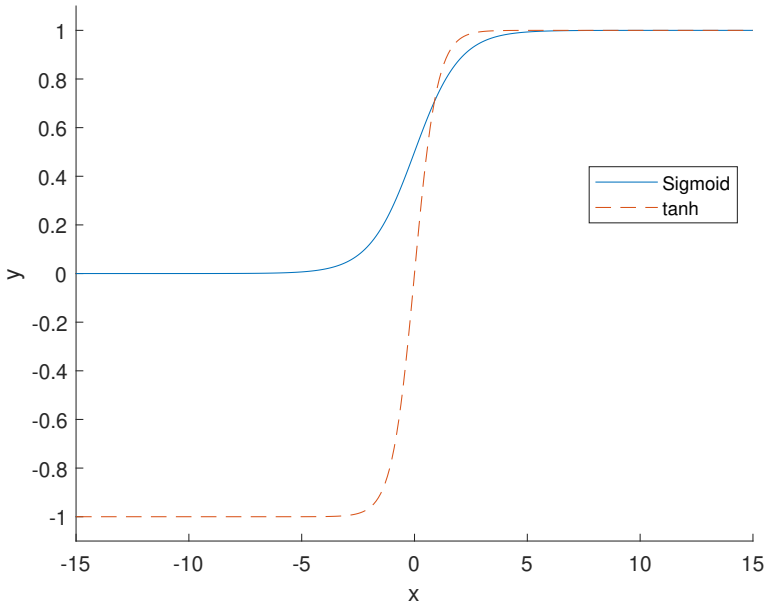



**Figure 4** – The activation functions Sigmoid and tanh.

Both these functions have their limitations as they saturate, which means that large values will output 1 and small values either output 0 or -1 depending on the function. Both functions are sensitive close to their midpoint of their input and outside of that area the sensitivity lessens quickly. Both this saturation and lack of sensitivity made it hard for learning algorithms to change the weights to improve the performance of the trained neural network. A result of hardware improvements allowed for deeper networks. This showed limitations in sigmoid and tanh such as not being able to train the deeper networks. This is due to a problem known as the vanishing gradient problem and was related to the use of these non-linear activation functions [11].

These issues continue into the 2000's, as the need for non-linearity for complex problem was counteracted by the difficulty for deep models to learn with the non-linear activation

functions. This was only solved in the early 2010's with the use of the Rectified linear unit (ReLU) activation function which is defined as

$$f(x) = max(0, x).$$

This piecewise linear function has become the standard for ANNs as it handles the saturation and sensitivity issues from both the Sigmoid and tanh functions, as well as dealing with the vanishing gradient problem. In Figure 5 the ReLU function is shown.

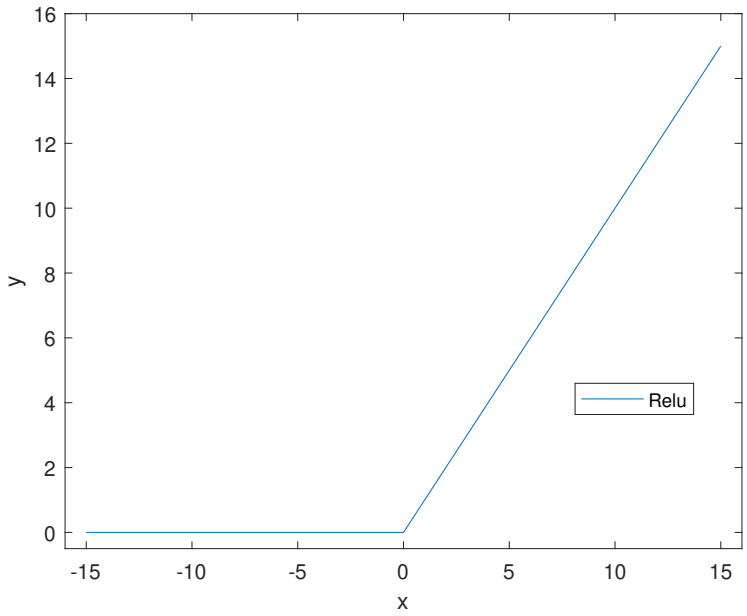



**Figure 5** – The activation function rectified linear unit.

The ReLU activation function to this day keeps being the standard for ANN's as it handles the above problems and simplifies operations as it only requires a comparison.

For classification problems, the last layer's activation function is often chosen to be the softmax function. This is due to the output can be interpreted as probability distribution and such give an indication of how certain the neural network is with the classification [12].

### 2.1.3 Loss function and optimizing

With the network defined with its weights, activation functions and biases it still need to be trained to solve different problems. This training is an optimization problem where the network works as a function $f(x)$, which takes input x and then outputs y, which then can be compared against already derived outputs y. These comparisons are done in different way, but it is defined as a loss function which is minimized during training.

In regression problems the mean square error is a possible loss function, but in more complex problems like categorical labelling more complex methods are needed. One way of doing this is by using cross-entropy. Cross-entropy comes from the field of information theory and is used to estimate the difference between an estimated and predicted possibility distribution.

This can then be used as a loss function as the difference between the predicted possibility distribution and the training dataset's possibility distribution can be minimized [11].

With the loss function defined, the training of the neural network becomes an optimization problem and therefore requires an optimizer. One optimizer strategy is using the stochastic gradient descent algorithm. The algorithm estimates the error gradient for the current state of the neural network using the training dataset. It then uses an error backpropagation algorithm to update the weights in the model in accordance with the step size, or as it is known in ANN, the learning rate. There are more complex versions of the stochastic gradient descent algorithm that can adapt the learning rate and one of these is the Adam algorithm [11, 13].

## 2.2 Evaluation metrics

To evaluate neural network models, multiple metrics are used to capture different features of the network. These metrics are based on the prediction made by the model compared against the actual classification of the data. In a multiclass classification problem, each of these comparisons are made for each class, which divides the predictions into four cases.

- **True Positive (TP):**

  The number of correct predictions of the specific class.

- **False Positive (FP):**

  The number of predictions that incorrectly classifies as the specific class.

- **False Negative (FN):**

  The number of predictions that classifies the specific class into another class.

- **True Negative (TN):**

  The number of predictions that correctly classifies another class into another class.

From these four cases the metrics precision, sensitivity, specificity, accuracy and the F-score. The precision is calculated as following for each of the classes,

$$Precision = \frac{TP}{TP + FP}.$$

The sensitivity is calculated as follows,

$$Sensitivity = \frac{TP}{TP + FN}.$$

The specificity is calculated as follows,

$$Specificity = \frac{TN}{TN + FP}$$

In multiclass classification problems the accuracy can be somewhat misleading, but it can give a measurement of how likely the network model places an individual correctly without

considering the difference between the classes. The accuracy can be calculated as follows,

$$Accuracy = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i TN_i + \sum_i FP_i + \sum_i FN_i}.$$

The last evaluation metric is the F-score, also called the $F_1 - score$, which balances the Sensitivity and the Precision metrics to give a more representative score for the overall performance of the network model

$$F - score = \frac{2 \cdot Sensitivity \cdot Precision}{Sensitivity + Precision}.$$

# 3 Method

## 3.1 Overview

With the goal being to both produce a functional neural network based around vHIT and analyzing the data, the project was split into 2 separate parts. The first part focused on understanding the data and processing it so it could be used as an input dataset to the neural networks in the second part of the project that focus on training and optimization of three different types of neural networks to the stated problem.

## 3.2 Participants

The participants that took part of this study were part of a larger project at University Hospital of Umeå. The VN group were recruited in the emergency room for an earlier acute vestibular syndrome study and their vHIT-results and recordings were saved for research purposes. The VN diagnosis was confirmed with caloric reflex test. The stroke patients were recruited at the stroke ward and emergency room at Umeå University Hospital and their vHIT-results were saved for future research. The control group were recruited from the hospital staff and relatives to inpatients and their vHIT-results were saved for future research. Informed consent was attained from all participants. The ethical Review Board of the Umeå University approved the study (2014/284-31M (2019-02881)), and it was performed in accordance with guidelines of the Declaration of Helsinki. The demographic data for the participants are presented in Table 1

**Table 1** – A table summarizing the participants in the study with total amount, sex, and mean age with standard deviation.

| | **Control** | **Stroke** | ***VN*** |
|---|---|---|---|
| **Number of participants** | 91 | 73 | 54 |
| **Women/Men** | 44/47 | 29/44 | 25/29 |
| **Age** | 55(18) | 72 (14) | 68(12) |

## 3.3 Data pre-processing

The vHIT-data was acquired from University Hospital of Umeå using the VHIT Ulmer (SYNAPSYS SA, France). The dataset consisted of xml-files containing the raw data from the vHIT for each of the participants. It consisted of the positional time series for each of the eyes and the head, as well as parameters of interest for the impulses performed in the head impulse test. The parameters extracted for further analysis were the VOR-gain and the Early Saccade Index (ESI), where the ESI is an index indicating how much of the impulse was affected by a saccade.

### 3.3.1 Filtering

Each of the individual's data was extracted from the VHIT Ulmer before the pre-processing was performed in Python 3.8.5. As an initial filter for the dataset, all the individuals with missing data for one of the vertebral canals were removed. This was done to create a more homogeneous dataset for the neural networks and to simplify the process. From these remaining individuals further filtering was done by checking that the positional time series was within expected limits. The positional series have the unit degrees, and the filter was set to filter out any impulses that had time series containing values larger than 500 as they

are just errors. The last filtering step were to discard the impulse if the head moved less than 2 degrees. With the filtering process completed this left the dataset with 37 VN cases, 46 stroke cases and 76 control cases. This meant that 17 VN cases, 27 stroke cases and 15 control cases were discarded in the filtering process.

With the filtering of individuals and impulses done comes the data processing, which consisted of averaging the positional series for all the vestibular canals and as well averaging the parameters of interest. For the positional series this was done for the head by averaging all the impulses for a specific channel and for the eyes an initial average of the right and the left eye was done before taking the average for impulses. As there are 6 vertebral canals this in turn results in 12 averaged positional series where 6 represent the head and 6 represent the eyes. During the averaging process there is a loss of variation for the time series. To capture the variations that occur during singular impulses, a standard deviation was calculated for the averaged series. This resulted in 12 standard deviation series.

The same averaging process was done for the parameters VOR-gain, but for the EIS parameter some extra operations were needed. When the no saccade was noticed by the software it sets the value to N/A, which meant that just averaging would cause problems. To handle this, it was decided to replace each N/A with 0 in the dataset, to indicate that no saccade was present and such decreasing the average. With these parameters averaged and extracted from the software, an attempt to capture variations in eye positional time series in a parameter was done. It was done by taking an average of the standard deviation series for the eye movement, and such having a third parameter to analyze for each vertebral canal. This last parameter was labelled VAR, as it correlates to the variance and to have a similar label as the other parameters.

### 3.3.2 Group comparison

For further analysis of the network's ability to classify the different participants, MATLAB R2021b (MathWorks,USA) was used for group comparison. Firstly, the group means of the parameters were compared, to see if any significant differences could be observed between the VN group, stroke group and control group. This was done by performing a one-way ANOVA test, with $\alpha = 0.05$, on each of the parameters in each channel, followed by a pairwise posthoc using the Tukey's honestly significant difference criterion.

To get a more visual comparison of the position vectors for both the eyes and the head, an average was taken between all participants for each of the groups resulting in 6 mean curves for the head and 6 mean curves for the eyes. For each of the vestibular canals a plot was produced containing the mean curves for head and eye movement for all the groups. In these plots the standard deviation was added as a shaded area for each of the groups.

### 3.3.3 Normalization

In the vHIT the head's movement will for most cases be larger than the eyes movement which caused a small issue when using these position vectors as input for a neural network. This was due to when the weights in the networks were defined, some feature with a larger scale might get values higher than others without any specific reason. To counter this, each vector was normalized based on the largest value in each channel for all participants which resulted in values being within the range of -1 to 1.

## 3.4 Parameter optimization using Tensorflow

For the construction of the neural networks in this project Tensorflow 2.6.0 (Google Brain Team, USA) was used and specifically the sequential Keras model [12, 14]. Tensorflow is an open-source framework developed by Google researchers to run machine learning, deep learning and more. It is used in wide range of fields, such as healthcare, language filters, fraud detection and many more [15]. One other reason for using this was that the official documentation is thorough with a lot of examples making the implementation easier. There is also a lot of unofficial tutorials online which give further examples.

Tensorflow allowed to use a hypertuner which makes it simpler to test the effect of different hyperparameters such as number of hidden layers, learning rate, and number of nodes in each layer [16]. It also has an interaction with CUDA and cuDNN which allowed the use of a graphic card for the training process, which should decrease the training time compared to running on a CPU.

# 4 Implementation

## 4.1 Networks

Networks with three different complexities were created for the project. The first network was set up to maximize the amount of data and therefore used the movement data from the vHIT, including all the different vertebral canals, and their standard deviations vectors. The second network was simplified by focusing on lateral channels in hope to make it easier for the network to learn, while still capturing the differences that might occur on the left and right side. The input was as in the first network the movement data from the vHIT, but it only uses the lateral channels movement data and its standard deviations vectors. The third and last network was even simpler by using the parameters VOR-gain, ESI and VAR as input for the different vestibular canals. For the rest of the report the first network will be labelled as the complex design-all network, the second network as the complex design-lateral network and the last network as the parameter design network.

### 4.1.1 Data pre-processing

All networks use a similar structure for the input, and it only differs in the type of data stored in it. The different data used were restructured into a 3d matrix, where one of the dimensions represent the participant, the second dimension either specifies the vertebral canal for the parameter network or the specific feature in the movement data networks. These features in the more complex networks were either a head movement vector, an eye movement vector, or a standard deviation vector for either of the movements.

For the complex design-all network there were 24 features, the head movements for the 6 channels, the eye movement for the 6 channel and then the 12 vectors representing the standard deviation for each of the movement vectors. An illustration of the input can be seen in Figure 6.

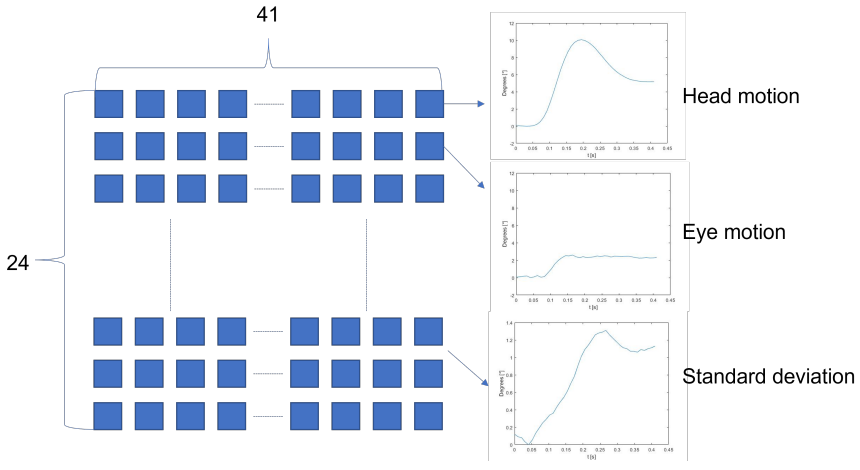



**Figure 6** – An example of the input to the complex design-all.

The complex design-lateral network has 8 features, the head movements of the 2 lateral channels, the eye movement for the 2 lateral channels and lastly the 4 vectors representing the standard deviation for each of the movement vectors. An illustration of the input can be seen in Figure 7.

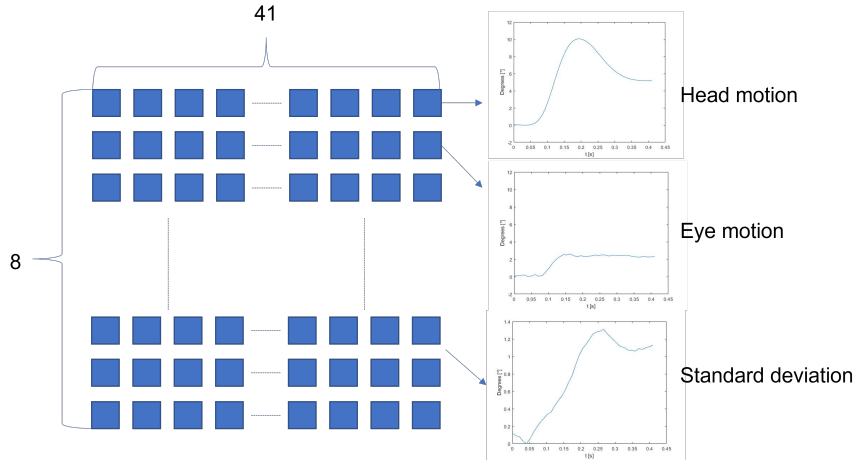



**Figure 7** – An example of the input to the complex design-lateral network.

The parameter design network has in the 3d matrix one dimension that represent the specific channel and then for each channel the 3 parameters, VOR-gain, ESI and VAR. An illustration of the input can be seen in Figure 8.

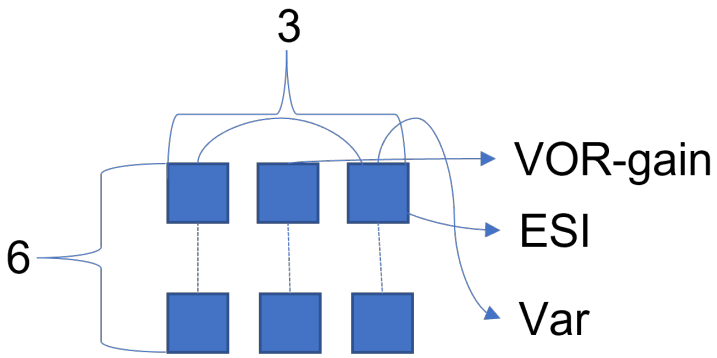



**Figure 8** – An example of the input to the parameter design network.

With this done the different groups were combined into one matrix and a label vector was produced containing the label for each participant in the main matrix. From these matrices 80% of the values were taken as training data and the rest was reserved as a test dataset for after the training. The labelling was done with one hot encoding which in turn means that the output of the network is a vector in $\mathbb{R}^3$ where the first node correlates to the control group, the second node to the stroke group and the last node correlates to the VN group.

### 4.1.2 Architecture

The networks defined above were built using a Keras sequential model, it is built layer by layer and has multiple different choices of layers. Fortunately, with the hypertuner these different combinations can easily be tested both when it comes to layer specific parameters and to the number of layers and type. The main layer that is the focus on in these networks were the Dense layer type which works as described in the 2.1.1 and the activation function was chosen to be ReLU due to the reasons discussed in 2.1.3. The initial layer takes in the

input in its 2d format and then is followed by a flatten functions that transforms the matrix to a 1d vector. The Hidden layers consists of multiple dense layers and the number of nodes for each of these and the input layer were optimized using the hypertuner. The output layer was set to three nodes to capture the three class classification problem and used the softmax activation function to give a probability distribution as its output.

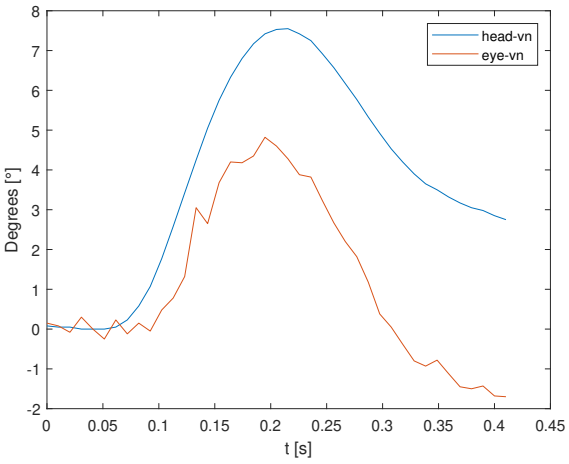
For the tuning process the number of layers was limited to be between 2 to 5 layers to not create a too complex network for our limited dataset. The number of nodes for the hidden layers and the input layers were limited to the range 5-19 during the tuning process. In the range the step size was set to 2, which meant that only half of the values in the range was. In a similar fashion the learning rate was limited to be between 1e-3 and 1e-4. To simplify the tuning the optimizer was locked to the popular Adam optimizer and finally the loss function was set to be categorical cross entropy. The resulting network from the hypertuning was then trained using the training data and tested against the test data.

# 5 Results

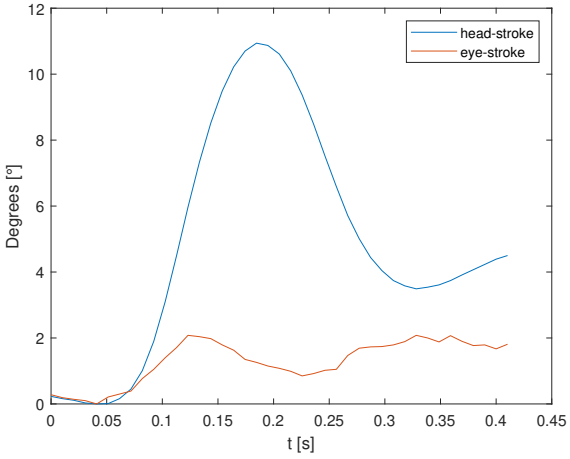
The results are split up into five sections, the first illustrating examples of time series data for the different groups, the second being a group comparison, and the last three containing the performance results of the networks.

## 5.1 Data example

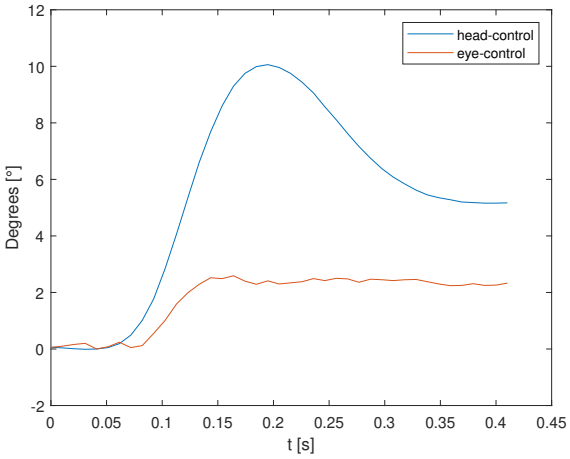
An example of the positional series for the eyes and the head for one participant from each of the groups can be seen in Figure 9a, 9b and 9c.




**(a)** An example of motion data for one VN participant and with the VOR-gain 0.22.



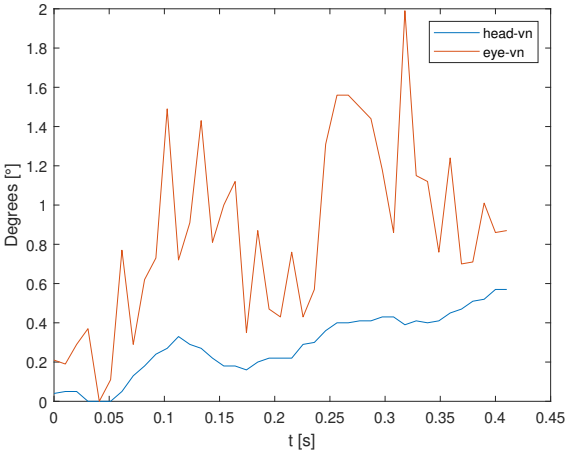

**(b)** An example of motion data for one stroke participant and with the VOR-gain 0.82.



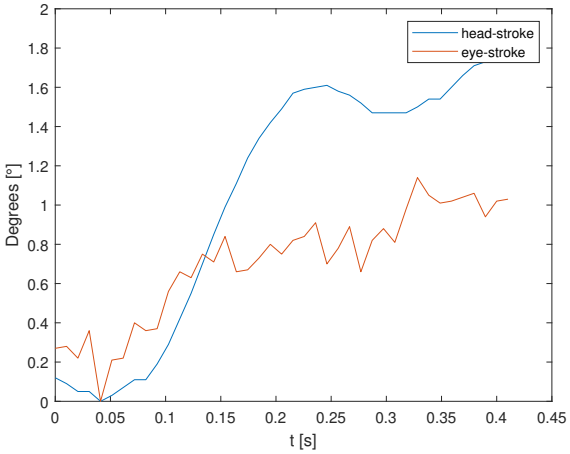

**(c)** An example of motion data for one control participant and with the VOR-gain 0.66.

**Figure 9** – Examples of motion data for a VN participant (upper left), a stroke participant (upper right) and a control participant (bottom).

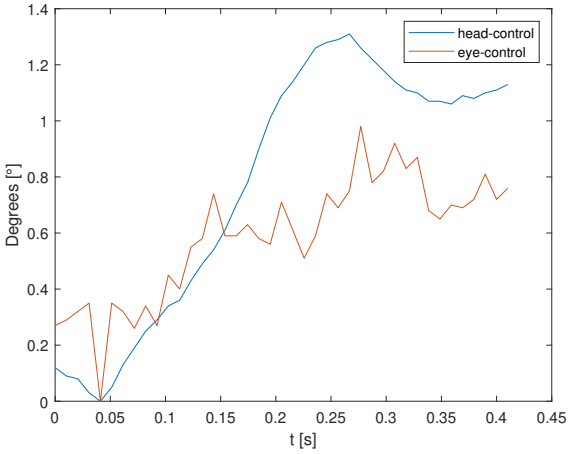
Further an example of standard deviation data for one participant from each of the groups can be seen in the Figures 10a, 10b and 10c.




**(a)** An example of standard deviation data for one VN subject and with the VOR-gain 0.22.




**(b)** An example of standard deviation data for one stroke subject and with the VOR-gain 0.82.




**(c)** An example of standard deviation data for one control subject and with the VOR-gain 0.66.

**Figure 10** – Examples of standard deviation data for a VN participant (upper left), a stroke participant (upper right) and a control participant (bottom).

## 5.2 Group comparison

This section is divided into two parts. The first part is a comparison between the parameters, and the second part is a comparison of the eye movement between the different canals

### 5.2.1 Group differences in gain and saccade parameters

The parameter comparison in Table 2 illustrates the mean of the different parameters for the left vertebral canals. For the left posterior canal, the VOR-gain was significantly different for the control group compared to both stroke and VN, while no differences were found between the Stroke and VN. For both the ESI and the VAR (for the same canal) no difference could be found, even though the ANOVA implied that there might be for the VAR parameter. This is due to the Tukey-Kramer pairwise comparison couldn't show any significant difference.

For the left anterior vertebral canal, the VOR-gain for the VN group was significantly different from both the control group and the stroke group. No difference could be found between stroke and control. For the ESI parameter, the VN group was significantly different from the two other groups but there was no difference between the control and stroke groups. Lastly,

the VAR parameter was significantly different for the VN group compared to the other two, while no difference could be found between the control and stroke groups.

For the left lateral vertebral canal, there was a significant difference in the VOR-gain between all groups, with the difference between the control group and the stroke group having a p-value close to being insignificant. For the ESI parameter, the VN group was significantly different from the two other groups and no difference could be found between the control and stroke groups. Lastly, the VAR parameter was significantly different for the VN group compared to the other two, while no difference could be found between the control and stroke groups.

**Table 2** – The parameters VOR-gain, Early Saccade Index (ESI) and the Variance mean parameter (VAR) for the left vertebral canals for the groups Control (*Ctrl*), Stroke (*Strk*), and Vestibular Neuritis (*VN*). The ANOVA column shows the p value for the parameter test and the Tukey-Kramer column shows any significant differences that might occur.

| | ***Ctrl*** | ***Strk*** | ***VN*** | **ANOVA (p value)** | **Tukey-Kramer (p value)** |
|---|---|---|---|---|---|
| **Left Posterior** | | | | | |
| VOR-gain | 0.74(15) | 0.65(18) | 0.62(17) | 0.0004 | Ctrl $\neq VN(0.0001)$<br>Ctrl $\neq Strk(0.0089)$ |
| ESI | 1.10(3.21) | 1.62(3.49) | 1.56(4.06) | 0.6531 | |
| VAR | 0.50(34) | 0.65(37) | 0.64(23) | 0.0297 | |
| **Left Anterior** | | | | | |
| VOR-gain | 0.81(16) | 0.80(17) | 0.38(20) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$ |
| ESI | 0.60(2.40) | 0.41(1.33) | 5.42(4.27) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$ |
| VAR | 0.45(21) | 0.50(26) | 0.80(36) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$ |
| **Left Lateral** | | | | | |
| VOR-gain | 0.80(13) | 0.68(22) | 0.25(36) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$<br>Ctrl $\neq Strk(0.0150)$ |
| ESI | 0.62(1.97) | 0.93(2.58) | 6.78(5.26) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$ |
| VAR | 0.41(27) | 0.59(38) | 1.27(69) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$ |

The parameter comparison in Table 3 illustrates the means of the different parameters for the right vertebral canals. For the right posterior vertebral canal, the VOR-gain for the VN group was significantly different from the other groups and that no difference could be found between the control and stroke groups. For the ESI, the stroke and VN group was

significantly different from each other, while no difference could be seen between control and the other groups. For the VAR, both stroke and control were significantly different from VN, but no difference could be found between stroke and control.

For the right anterior vertebral canal, the VOR-gain for the VN group was significantly different between from both the stroke and control group, while no difference was found between stroke and control. For the ESI, VN was significantly different from the other groups and no difference could be found between the control and stroke groups. For the VAR a significant difference was found between control and stroke group, while no difference could be found between VN and the other groups.

Lastly, the right lateral vertebral canal showed that the VOR-gain was significantly different for the VN group compared to the other groups. There was no significant difference between the control and stroke group. The ESI was found to be significantly different between VN and the other groups, while no difference could be found between control and stroke. The VAR was found to be significantly different for VN and stroke, but no difference could be found between stroke and the others.

**Table 3** – The parameters VOR-gain, Early Saccade Index (ESI) and the Variance mean parameter (VAR) for the right vertebral canals for the groups Control (*Ctrl*), Stroke (*Strk*), and Vestibular Neuritis (*VN*). The ANOVA column shows the p value for the parameter test and the Tukey-Kramer column shows any significant differences that might occur.

| | ***Ctrl*** | ***Strk*** | **VN** | **ANOVA (p value)** | **Tukey-Kramer (p value)** |
|---|---|---|---|---|---|
| **Right Posterior** | | | | | |
| VOR-gain | 0.75(13) | 0.66(15) | 0.60(17) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0022)$ |
| ESI | 0.94(2.96) | 1.29(2.80) | 2.91(3.87) | 0.0075 | Strk $\neq VN(0.0049)$ |
| VAR | 0.49(23) | 0.52(22) | 0.66(30) | 0.0019 | Ctrl $\neq VN(0.0260)$<br>Strk $\neq VN(0.0010)$ |
| **Right Anterior** | | | | | |
| VOR-gain | 0.82(16) | 0.77(20) | 0.50(20) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$ |
| ESI | 0.66(1.67) | 0.79(2.15) | 4.04(4.23) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$ |
| VAR | 0.48(29) | 0.63(47) | 0.60(17) | 0.0331 | Ctrl $\neq Strk(0.0395)$ |
| **Right Lateral** | | | | | |
| VOR-gain | 0.79(14) | 0.72(22) | 0.43(28) | 0.0000 | Ctrl $\neq VN(0.0000)$<br>Strk $\neq VN(0.0000)$ |
| ESI | 0.29(1.28) | 1.25(3.76) | 3.11(4.90) | 0.0001 | Ctrl $\neq VN(0.0245)$<br>Strk $\neq VN(0.0000)$ |
| VAR | 0.48(48) | 0.71(61) | 0.97(61) | 0.0001 | Ctrl $\neq VN(0.0000)$ |

### 5.2.2 Eye movement

In Figure 11 the mean curves for the eye´s motion are illustrated for each of the groups for the left canals, as well as a shaded area that represents the standard deviation for the mean curves. In Figure 11a the control and stroke group are closely related for the eye motion, while the eye motion for VN clearly differs for the left posterior canal. A similar pattern can be seen in Figure 11b, but here VN only differs slightly from the others on eye motion for the left anterior canal. Finally in Figure 11c, the eye motion mean curves for VN differ greatly from the other groups.

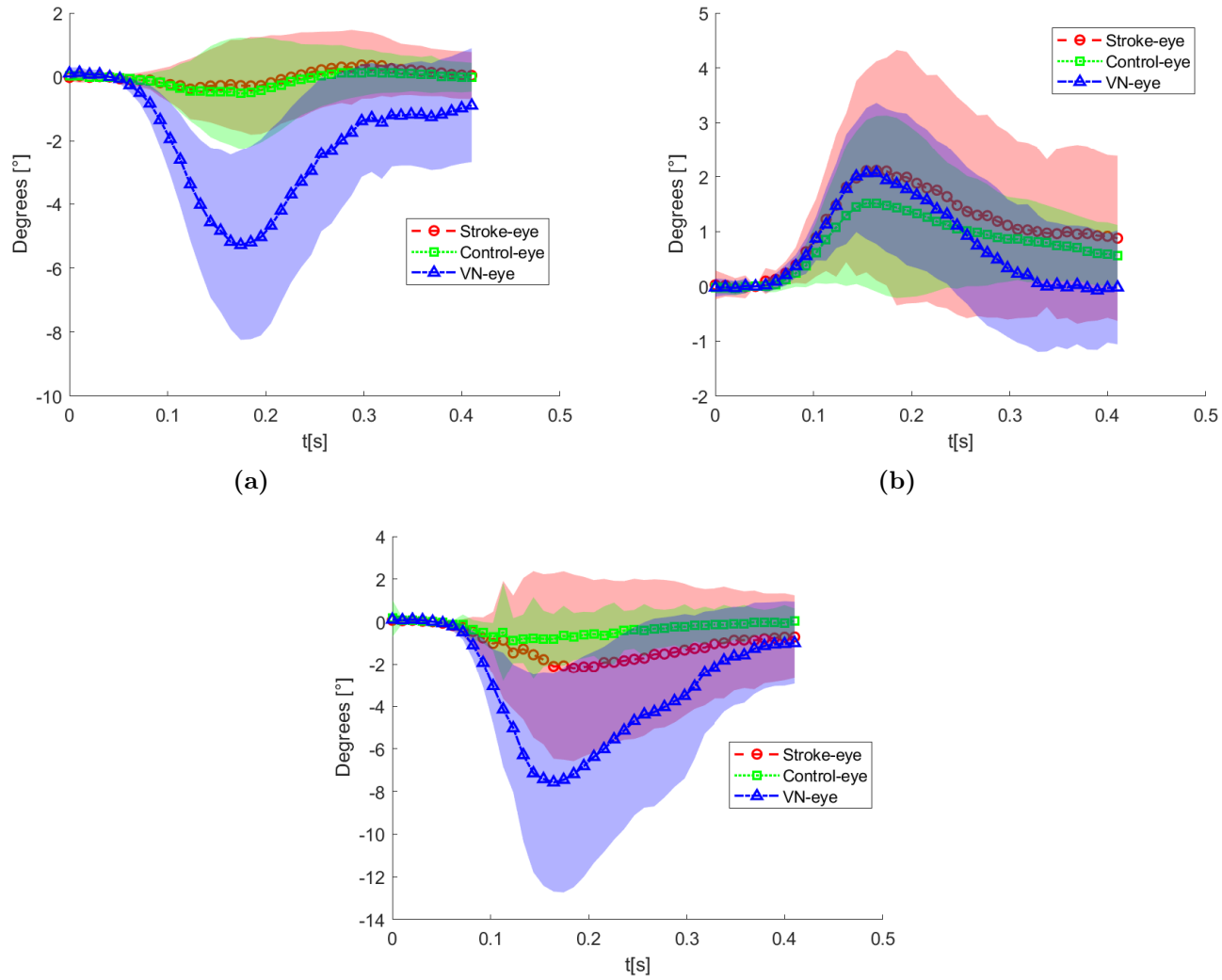



**Figure 11 – Eye motion during the vHIT:** The figures illustrate the mean eye motion for the different groups as well as the standard deviation for each of the groups as shaded areas. The red circle dotted lines and shaded area illustrate the stroke group, the blue triangle dotted lines and shaded area illustrate the VN group, and the green square dotted lines and shaded area illustrate the control group. Figure 11a illustrates the left posterior vertebral canal, Figure 11b illustrates the left anterior vertebral canal and Figure 11c illustrates the left lateral vertebral canal.

In Figure 12, the mean curves for the eye´s motion are illustrated for each of the groups for the right canals, as well as a shaded area that represents the standard deviation for the mean curves. In Figure 12a the control and stroke group are closely related for the eye motion, while the eye motion for VN slightly differs for the right posterior canal. A similar pattern can be seen in Figure 12b, but here VN only differs slightly from the others on eye motion for the right anterior canal. Finally in Figure 12c the eye motion mean curves for VN differ from the other groups.

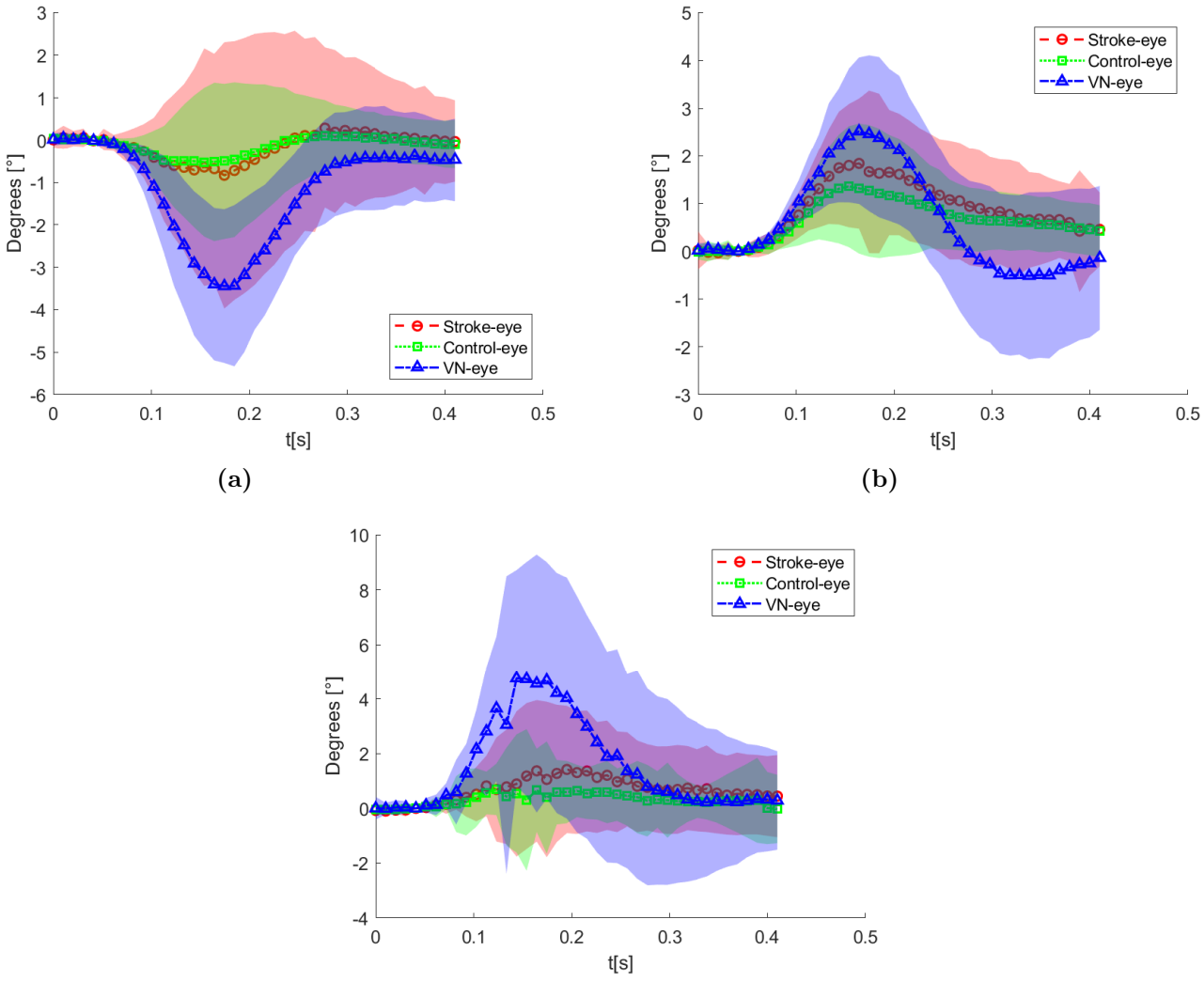



**Figure 12 – Eye motion during the vHIT:** The figures illustrate the mean eye motion for the different groups as well as the standard deviation for each of the groups as shaded areas. The red circle dotted lines and shaded area illustrate the stroke group, the blue triangle dotted lines and shaded area illustrate the VN group, and the green square dotted lines and shaded area illustrate the control group. Figure 12a illustrates the right posterior vertebral canal, Figure 12b illustrates the right anterior vertebral canal and Figure 12c illustrates the right lateral vertebral canal.

## 5.3 Network 1

The hypertuning process resulted in the network that follows the structure as described in Table 4.

**Table 4**

| **Layers:** | | |
|---|---|---|
| | **Nodes** | **Activation function** |
| **Input Layer: Dense** | 17 | ReLU |
| **Flatten Layer:** | | |
| **Hidden Layer 1: Dense** | 19 | ReLU |
| **Hidden Layer 2: Dense** | 19 | ReLU |
| **Output Layer: Dense** | 3 | softmax |
| **Compiling** | | |
| **Learning rate:** | 5.9 e-3 | |
| **Loss function:** | Categorical cross entropy | |
| **Optimizer:** | Adam | |

This neural network model was tested on the test dataset and this resulted in the confusion chart illustrated in Figure 13. The network performs with 77.8 % accuracy for Control group, 55.6 % accuracy for the stroke groups and 80.0 % accuracy for VN group.

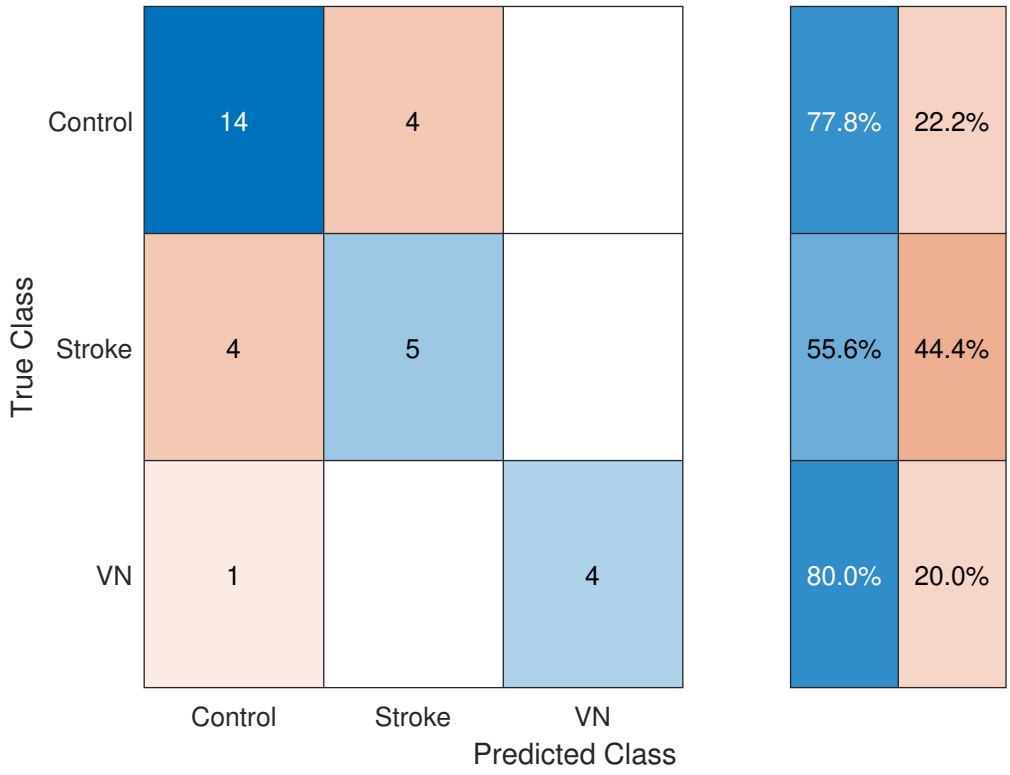



**Figure 13** – The left illustration shows Confusion chart for the complex design-all network. The right illustration shows the sensitivity for the different classes in the left column and the right column gives a percentage of misclassification for those classes.

From this confusion chart evaluation metrics were calculated which can be seen in Table 5.

**Table 5** – All channels Network performance table.

| Attributes | Control | Stroke | VN |
|---|---|---|---|
| **True Positive** | 14 | 5 | 4 |
| **False Positive** | 5 | 4 | 0 |
| **False Negative** | 4 | 4 | 1 |
| **True Negative** | 9 | 19 | 27 |
| **Precision** | 0.74 | 0.56 | 1.00 |
| **Sensitivity** | 0.78 | 0.56 | 0.80 |
| **Specificity** | 0.64 | 0.83 | 1.00 |
| **Accuracy** | 0.72 | 0.72 | 0.72 |
| **F-measure** | 0.76 | 0.56 | 0.89 |

In Figure 14 the output nodes activation is illustrated compared to the correct classification. With the information from Figure 13 and the node information, the network classified the control group correctly with a few exceptions. For the stroke classification it mostly confused the correct classification with control and not so much with VN. Lastly, the VN node was clearly distinct between the groups with only one case where a VN was classified as control.

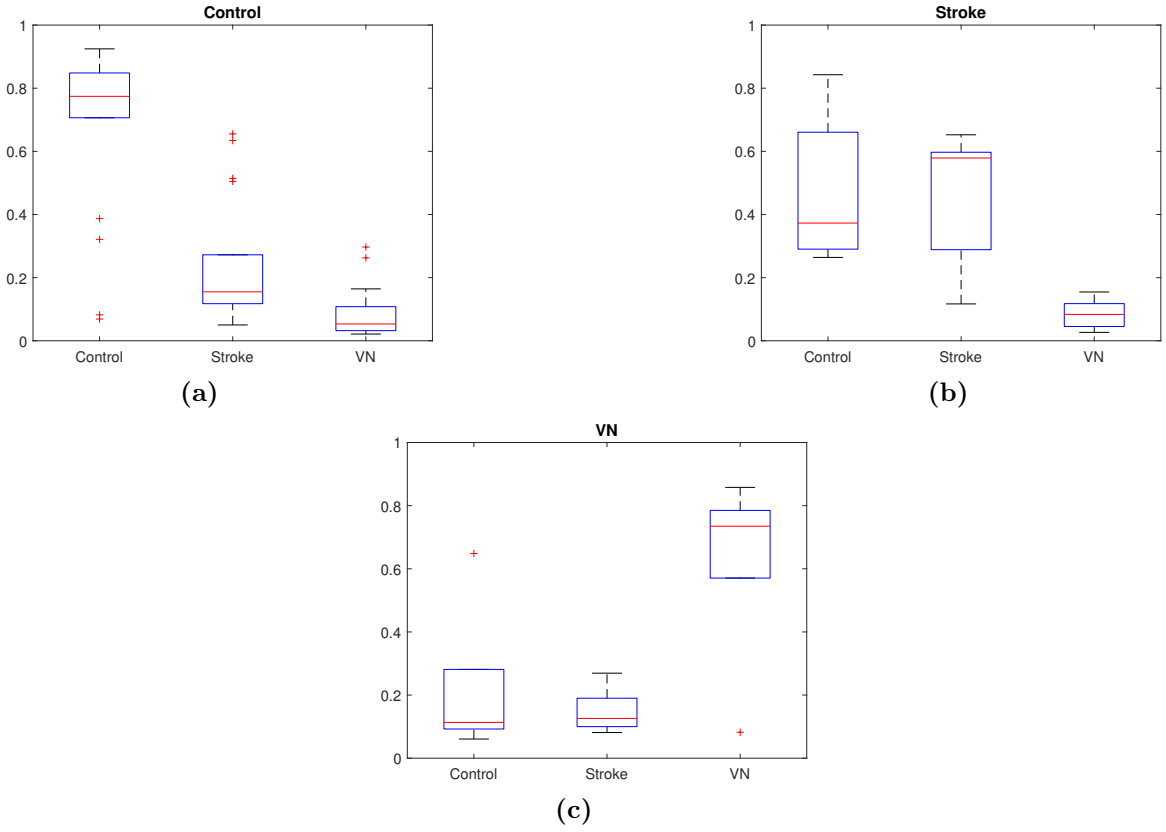



**Figure 14** – The node activation for each true classification for the complex design-all network using MATLAB's boxplot function. The left upper figure represent control node, the right upper figure represent stroke node and bottom figure represent the VN node

## 5.4 Network 2

The hypertuning process resulted in the network that follows the structure as described in Table 6.

**Table 6**

| **Layers:** | | |
|---|---|---|
| | **Nodes** | **Activation function** |
| **Input Layer: Dense** | 7 | ReLU |
| **Flatten Layer:** | | |
| **Hidden Layer 1: Dense** | 7 | ReLU |
| **Hidden Layer 2: Dense** | 19 | ReLU |
| **Hidden Layer 3: Dense** | 13 | ReLU |
| **Hidden Layer 4: Dense** | 17 | ReLU |
| **Output Layer: Dense** | 3 | softmax |
| **Compiling** | | |
| **Learning rate:** | 8.0 e-3 | |
| **Loss function:** | Categorical cross entropy | |
| **Optimizer:** | Adam | |

This neural network model was tested on the test dataset and this resulted in the confusion chart illustrated in Figure 15. The network performs with 100.0 % accuracy for Control group, 11.1 % accuracy for the stroke groups and 80.0 % accuracy for VN group.

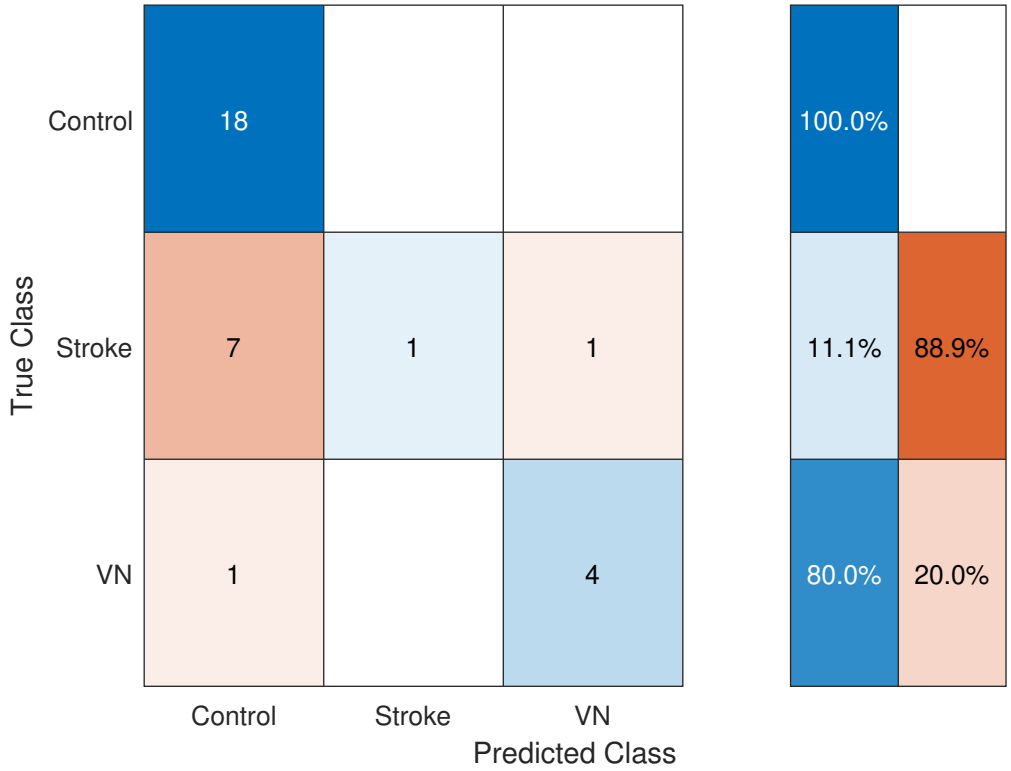



**Figure 15** – The left illustration shows Confusion chart for the complex design-lateral network. The right illustration shows the sensitivity for the different classes in the left column and the right column gives a percentage of misclassification for those classes.

From this confusion chart evaluation metrics were calculated which can be seen in Table 7.

**Table 7** – Lateral channels Network performance table.

| Attributes | Control | Stroke | VN |
|---|---|---|---|
| **True Positive** | 18 | 1 | 4 |
| **False Positive** | 8 | 0 | 1 |
| **False Negative** | 0 | 8 | 1 |
| **True Negative** | 6 | 23 | 26 |
| **Precision** | 0.69 | 1.00 | 0.80 |
| **Sensitivity** | 1.00 | 0.11 | 0.80 |
| **Specificity** | 0.43 | 0.74 | 0.80 |
| **Accuracy** | 0.72 | 0.72 | 0.72 |
| **F-measure** | 0.82 | 0.20 | 0.80 |

In Figure 16 the output nodes activation is illustrated compared to the correct classification. With the information from Figure 15 and the node information, the network classifies the control group correctly with a few exceptions. For the stroke classification it clearly classifies a most of the stroke group as control instead of stroke. Finally, it shows only one VN case was incorrectly classified as a control case.

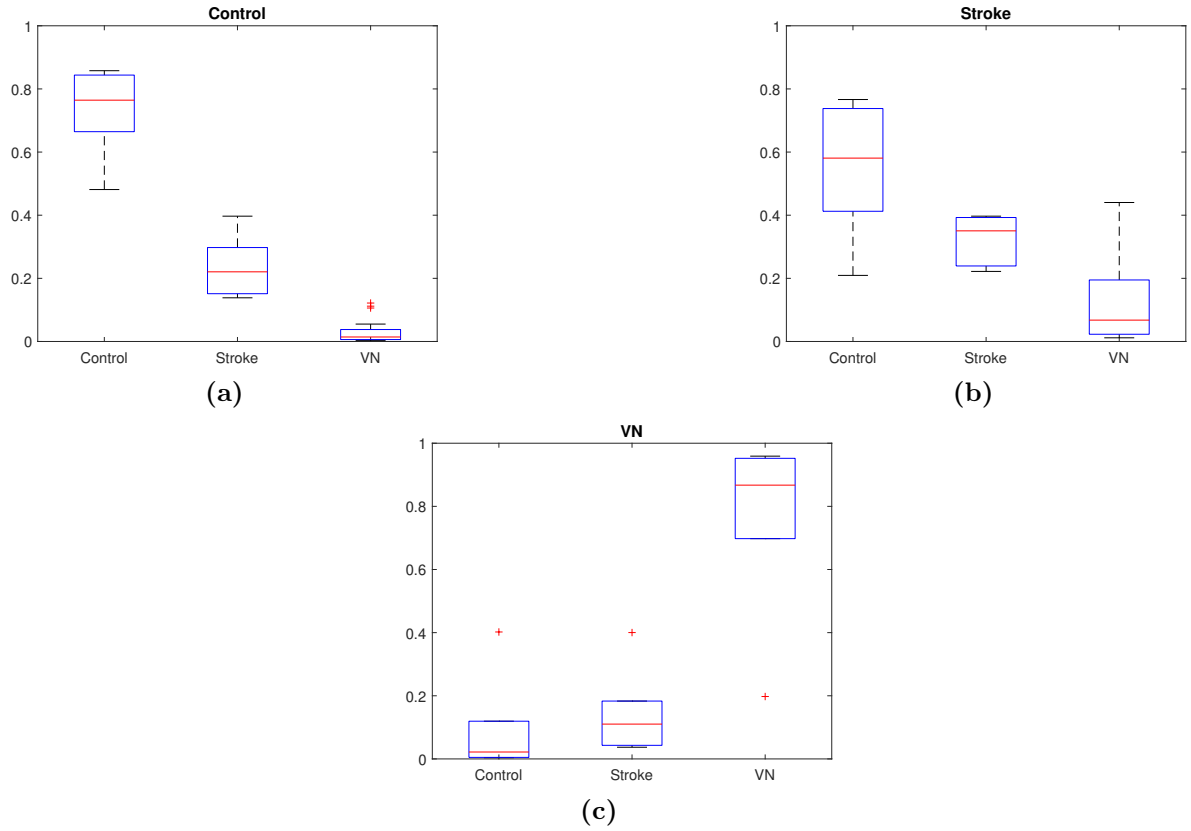



**Figure 16** – The node activation for each true classification for the complex design-lateral network using MATLAB's boxplot function. The left upper figure represent control node, the right upper figure represent stroke node and bottom figure represent the VN node

## 5.5 Network 3

The hypertuning process resulted in the network that follows the structure as described in Table 8.

**Table 8**

| **Layers:** | | |
|---|---|---|
| | **Nodes** | **Activation function** |
| **Input Layer: Dense** | 19 | ReLU |
| **Flatten Layer:** | | |
| **Hidden Layer 1: Dense** | 19 | ReLU |
| **Hidden Layer 2: Dense** | 17 | ReLU |
| **Output Layer: Dense** | 3 | softmax |
| **Compiling** | | |
| **Learning rate:** | 7.7 e-3 | |
| **Loss function:** | Categorical cross entropy | |
| **Optimizer:** | Adam | |

This neural network model was tested on the test dataset and this resulted in the confusion chart illustrated in Figure 17. The network performs with 88.9 % accuracy for Control group, 44.4 % accuracy for the stroke groups and 100 % accuracy for VN group.

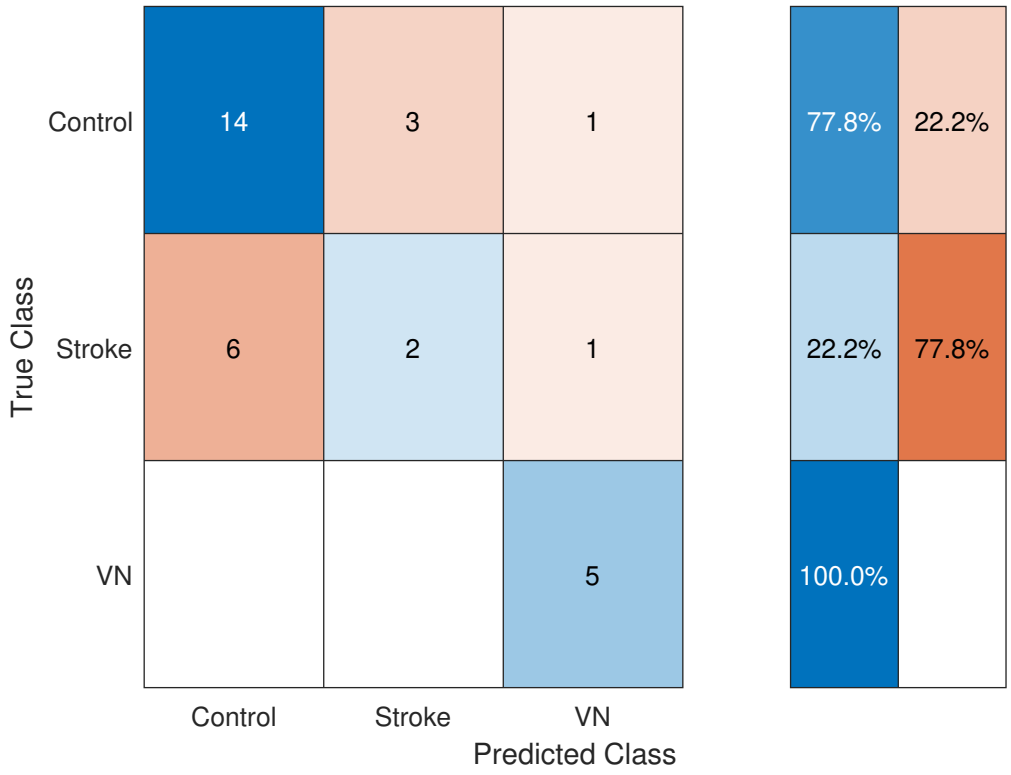



**Figure 17** – The left illustration shows Confusion chart for the parameter design network. The right illustration shows the sensitivity for the different classes in the left column and the right column gives a percentage of misclassification for those classes.

From this confusion chart evaluation metrics were calculated which can be seen in Table 9.

**Table 9** – Parameter Network performance table.

| Attributes | Control | Stroke | VN |
|---|---|---|---|
| **True Positive** | 14 | 2 | 5 |
| **False Positive** | 6 | 3 | 2 |
| **False Negative** | 4 | 7 | 0 |
| **True Negative** | 8 | 20 | 25 |
| **Precision** | 0.70 | 0.40 | 0.71 |
| **Sensitivity** | 0.78 | 0.22 | 1.00 |
| **Specificity** | 0.57 | 0.87 | 0.93 |
| **Accuracy** | 0.66 | 0.66 | 0.66 |
| **F-measure** | 0.74 | 0.29 | 0.83 |

In Figure 18 the output nodes activation is illustrated compared to the correct classification. With the information from Figure 15 and the node information the network doesn't clearly distinct between stroke and control. For the stroke classification the network can't distinct between stroke and control and it even classifies most stroke cases as control. Finally, that all VN cases was distinctly classified correctly.

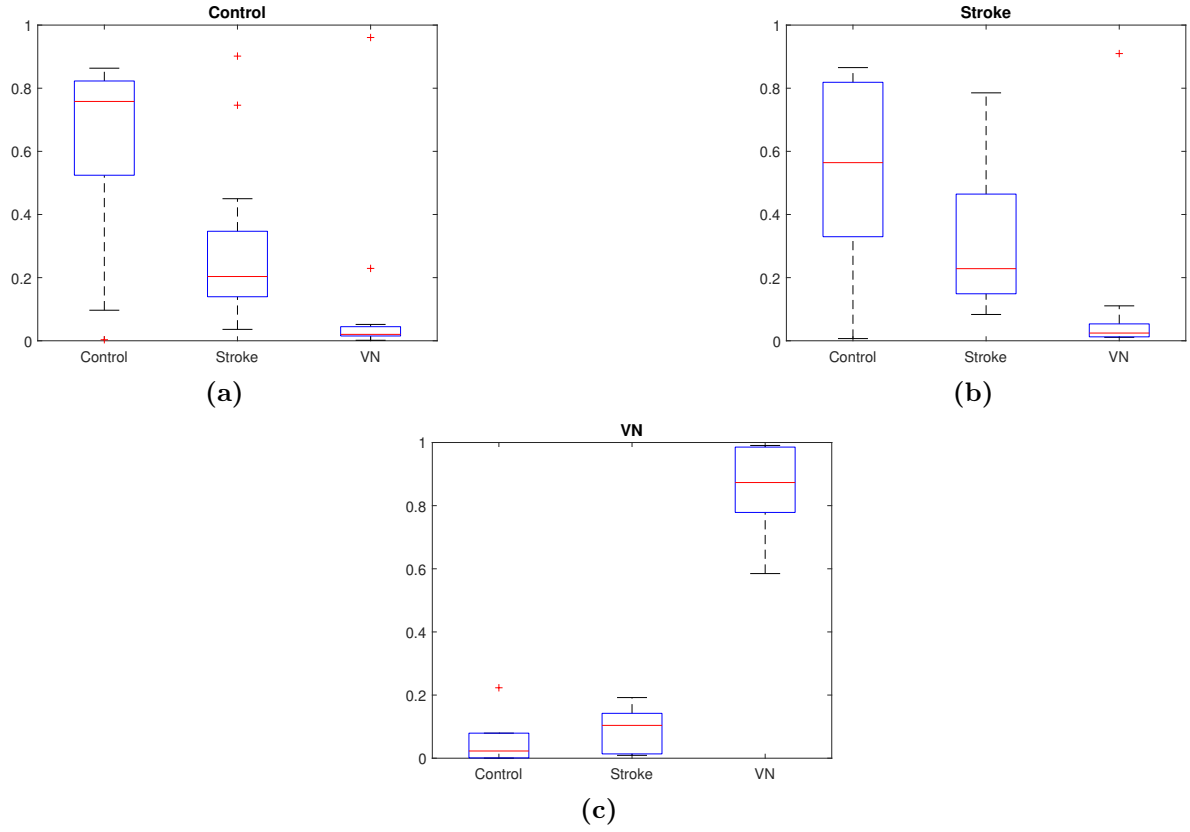



**Figure 18** – The node activation for each true classification for the complex parameter design network using MATLAB's boxplot function. The left upper figure represent control node, the right upper figure represent stroke node and bottom figure represent the VN node

# 6 Discussion

## 6.1 Dataset

The SYNAPSYS hardware and software records the motion of the eyes and the head movement, and the parameters were all based on these measurements. All the data used in this project was collected using the same hardware setup. The groups differed from patients at the hospital to the control group which consisted of recruits from the hospital staff and inpatient relatives. The control group was the largest group before filtering (as seen in Table 1) as participants in the control group were easier to collect, and had less issues with the test itself. The filtering process also removed some of the persons in the stroke and the VN group as they might have problems handling the full test and therefore only producing motion data for some canals. This created a difference in the sample sizes for the different groups which could have affected the performance of the selected neural networks. There was also a difference in age between the different groups as the control group was notably younger compared to the other two. This is something to consider as a possible limitation of this analysis, especially in of the research by Hansson and Salzer that small decreases of the vestibular ocular reflex gain might occur as individuals gets older [17].

As seen in Figures 9 and 10 it is clear that the differences between individuals in the different groups can be quite distinct, while analyzing individual time series.

The parameters extracted from Table 2 and Table 3 show that the VOR-gain significantly differs for the VN group compared to the others for most of the canals. The stroke group's parameters often overlap with the control group, but there were two canals in the dataset where there was a significant difference between the two. This significant difference was unfortunately not enough as these canals overlapped with VN, which limits the ability to discern between stroke and VN. This seems to agree with what Calic and Guler have found as they both added additional criteria to discern between stroke and VN [6, 5].

Most of the values for the ESI were around zero due to how most cases do not get an ESI value and therefore are set to zero, as previously discussed in the method. The impulses that have a value were often clearly larger than 0 and this leads to a large standard deviation. The group that most often had a nonzero ESI was the VN group and it was significantly different from the others in all canals except the left posterior vertebral canal.

Lastly, the parameter VAR, which I calculated myself from the eye motion data to try to capture the variances that might be lost during the averaging process in the pre-processing. This behaves similarly as the ESI parameter where a significant difference was found mostly between the VN group and the other groups with only a few exceptions.

From Figure 11 and Figure 12 the differences between specifically VN and the other groups were very distinct at least by visual comparison. This was expected as VN often has drastic deficiencies in functionality on either the left or right vestibular canals. This in turn will affect the vestibular ocular reflex which causes the eye to follow the motion of the head during the vHIT procedure. Between the stroke group and the control group the difference were not as distinct. This was one of the reasons to attempt using neural networks to differentiate between them.

## 6.2 Networks

In the following section I will discuss the results of the different networks. Firstly, the complex design-all network performed the best when compared to the other networks. From the confusion chart in Figure 13 it shows that the network could handle both the VN and control cases quite well. No cases were incorrectly classified as VN, while multiple cases of stroke and one of VN were classified as control. The stroke classification had a worse performance since a lot of stroke cases were classified as control. For the stroke cases it was important to have a high sensitivity and (as seen in Table 5) it only reached approximately 56 % which would cause misclassification of multiple stroke cases. For the control group it was instead important to have a higher specificity as that means less cases being sent for further unnecessary examinations. For VN a high sensitivity is wanted to correctly classify each VN case and from Table 13 it is shown to reach a high sensitivity. In Figure 14 its shown how the output layer has been activated for each of the groups. It illustrates clearly that the networks can distinct between the VN and control group quite well as the nodes activation does not overlap for most cases. As the last layer uses softmax activation function it means that it gives out a probability of classification. This means that the network is certain in the classification for both control and VN. For stroke instead, there is a big overlap with the control group and such it was not as certain with its classification. This overlaps with the information from the confusion chart as most incorrectly stroke cases are classified as a control case.

Secondly, the complex design-lateral network performs worse than the first network. From the confusion chart in Figure 15 it shows that the network correctly classified all the control cases, but it incorrectly classified almost all the stroke cases into the control group. It handles the VN group in a similar fashion as the first network, with the only difference being that it classifies one stroke case as a VN case. In summary, this network only performs better than the first network in classifying control group cases correctly, but it also gave more false positives and therefore misses critical diagnosis. One interesting point was that it does performed well with diagnosing VN cases which means that just using one canal from each side gives enough data for the network to distinct them from the other cases. As VN affects the vestibulocochlear nerve on one side of the head, using only one canal from each side, highlights the decreased function enough for the classification to work. For the stroke cases instead the opposite occurs as a stroke can happen in different places in the brain. This means that removing the information from the other canals gives a less complete picture of the VOR functionality which may be affected by the stroke. In Figure 16 it shows how the output layer has been activated for each of the groups and in this case, the same distinction is clear between VN and control while the stroke group overlaps with both the control group and VN giving less secure classifications. Comparing this figure with the confusion chart makes it clear that the network mostly confused the stroke group with the control group.

The last network performs better than the more complex design-lateral network but handles stroke cases worse than the complex design-all network. From the confusion chart in Figure 17 it shows that it correctly classifies the same percentage of control cases as the first network and outperforms the first network while classifying the VN cases. For the stroke cases it instead performed just slightly better than the complex design-lateral network and classified most stroke cases into the control group. This seemed to correlate with the information from the parameter comparison as no significant difference could be found for most of the

parameters between stroke and control while VN often differed from both groups. In Figure 18 it shows how the output layer has been activated for each of the groups and that the network is confident in its classification for VN while the other two nodes have larger overlaps. Comparing this with the confusion chart shows how it clearly confuses stroke and control and that this mostly happens towards control.

In summary, the networks all differ greatly when it comes to the classification of the stroke cases, with the first network correctly classifying with the most cases while both the second and the third network both incorrectly classified most of the stroke cases. The parameter design network performed better than the complex design-lateral network, so it showed how a simpler setup have the possibility of outperforming a more complex network as can be seen from the Figures 15 and 17. All the networks also followed a very simple structure with dense layers and only a few hidden layers, so the biggest difference was the input provided to the networks. The parameter excelled on VN classification because of its great dependence on the VOR-gain while the complex design-all network handles the stroke cases better as it had an input which gave a broader description of an individual's VOR compared to the lateral canal network.

## 6.3 Clinical Application

With the technical discussion done I want to bring it back to medical application. The goal for this project was to research the possibility to use deep learning to correctly classify a patient from the results of a vHIT and for this to be used as a diagnostic tool in for example an emergency room. As the results stands now, the sensitivity is too low for the detection of a stroke for a patient, but these networks do show promising tendencies for these kind of solutions in the future. The sample size for these networks were both unbalanced and quite small for the realm of deep learning and even with that they reached a decently high classification for both control and VN. The complex design-lateral network performed the best on average of the three networks, with correctly classifying around 80 % of both VN and control and by classifying around 60 % of the stroke cases. With further collection of data from the different groups and by further improving the network themselves, a higher performance could be reached. It is also possible to not have to limit the data collection to specifically SYNAPSYS AB hardware as the only requirement is to be able to record the positional data in a similar structure.

If this was done, it could be used widely in emergency rooms as the process would be simple and non-intrusive. The vHIT would be performed and recorded and the neural network would give out a diagnosis with the probability estimate on how sure it is.

## 6.4 Future work

This project has had a limited scope both due to time constraints and amount of data, but there are many improvements that can be done both to the networks and data itself.

Firstly, the dataset is unbalanced and quite small compared to dataset used in neural networks to train a model. The unbalanced nature in the data makes it so that it has very few cases to use as test data for the stroke and VN group and that limits the training. This is unfortunately also due to the procedure itself sometimes makes it so that subject in the stroke and VN group cannot produce complete result, due to things like a stiff neck or other limitations. This meant that the pre-processing filters out quite a few which further decreases

the dataset. This means that creating a balanced dataset is quite limiting, giving just above 100 total individuals when using the smallest group's sample size. Instead, I believe that gathering more data would benefit the performance the most, as then balancing the groups will not limit the network training process to a small dataset. The hardware used for the data collection exists in multiple hospitals and data collection should be able to be uniform thus creating the possibility to collect data from multiple sources and then combining them to a larger dataset.

For the networks there is a large potential for improvement with more time and experimenting. The networks in the present study are based around simple dense layers which receives input from all nodes of the previous layer. One other common layer that could be further tested with more time would be the convolution layer, which is commonly used to detect patterns in for example pictures. It uses filters to affect part of the input and therefore picking out specific features. This might be useful for our motion curves, detecting things like saccades or other discrepancies. This is often combined with pooling layers that is used to reduce dimensions of the feature map produced by the convolution layers. Networks using these layers combined with dense layers can have great performances, but it does create a more complex tuning process. With more time the current dataset might still be able to be used to create a model that performs better than the current versions.

The parameter network right now uses the VOR-gain which is the most common parameter from a vHIT, the ESI from the SYNAPSYS software and the VAR which is an average created from the standard deviation for the eye motion. From the comparison between the different groups, no clear distinction between the stroke and control groups could be statistically confirmed, which does not help the parameter network. If some other parameter could be produced either based on the saccades or some other feature, then this might make it possible to get a greater performance out of the parameter network as it already handles the VN case well.

# 7 Conclusion

In conclusion, this project has been a good first step in the possibility of using deep learning for diagnosing patients, using the non-invasive vHIT. It shows potential for further improvement as the motion data manages to produce effective classifications for VN and control cases with a small dataset. It is still lacking on the stroke cases, but with more data and further research this could be a first step on a road to diagnostic assistance using deep learning.

These models have shown promise for future application as decision support systems and especially the complex design-all network. If these networks are improved then it could hopefully not only save resources in the medical field as it is a simple method, but also catch time crucial diagnosis early which can then be giving the treatment they need.

# References

[1] Dao Nguyen, Insu Won, Kyusung Kim, and Jangwoo Kwon. A development of clinical decision support system for video head impulse test based on fuzzy inference system. *Journal of Sensors*, 2018:1–10, 2018.

[2] Thi Anh Dao Nguyen, Dae Young Kim, Sang Min Lee, Kyu Sung Kim, Seong Ro Lee, and Jang Woo Kwon. Recommendation application for video head impulse test based on fuzzy logic control. *Journal of Central South University*, 23(5):1208–1214, 2016.

[3] Anis Kacem, Zakia Hammal, Mohamed Daoudi, and Jeffrey Cohn. Detecting Depression Severity by Interpretable Representations of Motion Dynamics. In *IEEE FG Workshop, Face and Gesture Analysis for Health Informatics (FGAHI)*, Xi'an, China, May 2018.

[4] M. Ljunggren, J. Persson, and J. Salzer. Dizziness and the acute vestibular syndrome at the emergency department: A population-based descriptive study. *Eur Neurol*, 79(1-2):5–12, 2018.

[5] Zeljka Calic, Benjamin Nham, Andrew P. Bradshaw, Allison S. Young, Sonu Bhaskar, Mario D'Souza, Craig S. Anderson, Cecilia Cappelen-Smith, Dennis Cordato, and Miriam S. Welgampola. Separating posterior-circulation stroke from vestibular neuritis with quantitative vestibular testing. *Clinical Neurophysiology*, 131(8):2047–2055, 2020.

[6] A. Guler, F. Karbek Akarca, C. Eraslan, C. Tarhan, C. Bilgen, T. Kirazli, and N. Celebisoy. Clinical and video head impulse test in the diagnosis of posterior circulation stroke presenting as acute vestibular syndrome in the emergency department. *J Vestib Res*, 27(4):233–242, 2017.

[7] A. Schmid-Priscoveanu, A. Böhmer, H. Obzina, and D. Straumann. Caloric and search-coil head-impulse testing in patients after vestibular neuritis. *J Assoc Res Otolaryngol*, 2(1):72–8, 2001. 1438-7573 Schmid-Priscoveanu, A Böhmer, A Obzina, H Straumann, D Evaluation Study Journal Article Research Support, Non-U.S. Gov't 2001/09/08 J Assoc Res Otolaryngol. 2001 Mar;2(1):72-8. doi: https://doi.org/10.1007/s101620010060.

[8] Diagnosis using mri-scan. https://www.nhs.uk/conditions/stroke/diagnosis/. [Accessed 20 October 2021].

[9] Hearing and equilibrium | anatomy and physiology. https://courses.lumenlearning.com/nemcc-ap/chapter/special-senses-hearing-audition-and-balance/. [Accessed 7 October 2021].

[10] A. P. Landry, W. K. C. Ting, Z. Zador, A. Sadeghian, and M. D. Cusimano. Using artificial neural networks to identify patients with concussion and postconcussion syndrome based on antisaccades. *J Neurosurg*, pages 1–8, 2018.

[11] J. Brownlee. *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery, 2018.

[12] Keras documentation: Keras api reference. https://keras.io/api/. [Accessed 22 September 2021].

[13] Tensorflow adam optimizer. https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam. [Accessed 20 October 2021].

[14] Tensorflow core v2.6.0. https://www.tensorflow.org/api_docs/python/tf. [Accessed 22 September 2021].

[15] Tensorflow case studies. https://www.tensorflow.org/about/case-studies. [Accessed 22 September 2021].

[16] Tensorflow hypertuner. https://www.tensorflow.org/tutorials/keras/keras_tuner. [Accessed 22 September 2021].

[17] Anders Hansson and Jonatan Salzer. Normative video head impulse test data in subjects with and without vascular risk factors. *European Archives of Oto-Rhino-Laryngology*, 2020.