Comparing pain related EEG patterns in musculoskeletal pain, healthy participants, and experimental induced pain, using a machine learning algorithm

Biomedical Engineering | Group 10408
30th May 2025

Student project

AALBORG UNIVERSITY
STUDENT REPORT

Department of Health Science and Technology
www.hst.aau.dk

Title:
Comparing pain related EEG patterns in musculoskeletal pain, healthy participants, and experimental induced pain, using a machine learning algorithm.

Project topic:
Biomedical signal analysis and processing

Project period:
February - May 2025

Project group:
Group 10408

Participant:
Anders Kruse Laursen

Supervisor:
Sabata Gervasio

Pages total: 59

Date of completion:
30th May 2025

Abstract:

Introduction: Musculoskeletal (MSK) pain is the second leading cause of disability worldwide, affecting 20 % of the global population. The current methods of estimating MSK pain do not evaluate the psychological effects of pain when measuring or estimating the patient's pain sensation. *Aim:* The aim of this project was to develop a machine learning (ML) model, which analyses an objective measurement from the patient, and classifies it as either chronic, healthy or acute pain. While visualising where in the signal the pain features for each class are present. *Method:* The dataset was divided into segments, each including 5 seconds of electromyography (EEG) data, with the movement initiation aligned in all segments. A total of 1383 segments was used to train and test the developed Long Short-Term Memory (LSTM) Convolutional Neural Network (CNN). This included the implementation of heatmaps for temporal visualisation of the attention during the analysis. *Results:* The LSTM CNN had a general prediction accuracy of 54.55 % and a F1 score of 0.543. With heatmaps display the attention weights for solo segments, each class, and the general ML model. *Conclusion:* The system was able to predict the classes with an ok accuracy, while displaying the attention weights for the extracted features, thus further improvements are required for the healthy class's prediction accuracy and general prediction accuracy.

Dansk abstract

Introduktion: Muskuloskeletal (MSK) smerte er den anden største årsag til funktionsnedsættelse og påvirker 20 % af den globale befolkning. De nuværende metoder til estimering af MSK smerte tager ikke højde for de psykologiske påvirkninger smerte har på patienten. *Formål:* Formålet med dette projekt er at udvikle en machine learning (ML) model, der ved hjælp af en objektiv måling fra en forsøgsperson, kan klassificere om målingen er fra en akut eller kronisk smerte forsøgsperson, samtidig med at kunne visualisere hvor i signalet smertekarakteristika for hver gruppe var til stede. *Metode:* Datasættet blev opdelt i segmenter, som hver indeholder 5 sekunders elektroencefalografi (EEG) data, hvor at alle segmenter var justeret således at begyndelsestidspunktet for bevægelsen var ens i alle segmenterne. I alt blev 1383 segmenter gemt og brugt til at træne og teste den udviklede Long Short-Term Memory (LSTM) Convolutional Neural Network (CNN) model. Dette inkluderede derudover implementering af heatmaps til at visualisere hvor i segmenterne de forskellige features var udtrukket fra, som der var blevet brug til at klassificere dem. *Resultater:* LSTM CNN modellen opnåede en generel prædiktions nøjagtighed på 54,55 % og en F1 score på 0,543. Heatmaps blev anvendt til at vise vægtene af hvor vigtige de forskellige områder var for henholdsvis enkelte segmenter, for hver klasse og den overordnede ML model. *Konklusion:* Systemet var i stand til at klassificere med en ok nøjagtighed mellem akut of kronisk smerte, samtidig med at vise hvor i de forskellige segmenter eller klasser opmærksomheden var, under analysen. Systemet kræver dog en fremtidig forbedring, for at få en bedre prædiktions nøjagtighed af den raske forsøgspersons gruppe, hvilket potentielt også vil forbedre den overordnede prædiktions nøjagtighed af hele ML modellen.

Preface

This project was written by a university student from the fourth semester of the master's Biomedical Engineering and Informatics at Aalborg University. The project was written in the period from the 3rd of February 2025 to the 30th of May 2025. The project was written under the supervision of Sabata Gervasio.

Aalborg University, May 30, 2025

Anders Kruse Laursen
akla20@student.aau.dk

Reading instructions

Reference style

The Vancouver reference style was used throughout this project to cite other articles. In this style, references are indicated by numbers in square brackets, assigned in the order they appear in the text. E.g. the first citation is [1], the second is [2], etc. When referring to the article itself, the format follows this example: "The article by Author et al. [1] states that..", to clarify that the following information is drawn from that specific source. If a reference appears before a full stop, it means that the information in the specific sentence originates from that reference. If reference appears after the full stop, it means that the information in the section of the text, originates from this reference. Figure captions are located under the corresponding figure, while table captions are placed above the corresponding table.

Time plan

A time plan is provided after the conclusion. It outlines how time was allocated across the different phases of the project and describes the challenges encountered and how they were addressed during the time management for this project.

Nomenclature

Multiple abbreviations are used throughout the project, which are all listed here to provide an overview. Generally, the full word for abbreviations is only described once, after which only the abbreviation is used, e.g. Musculoskeletal (MSK). Hereafter, only the abbreviated word MSK is used. Table 1 illustrates all the abbreviations and the corresponding full words used throughout the project.

Table 1: *Illustrates the abbreviations and corresponding meaning used throughout the project.*

Abbreviation	Meaning
ACC	Anterior Cingulate Cortex
AIC	Anterior Insular Cortex
BOLD	Blood-Oxygen-Level-Dependent
CNS	Central Nervous System
CNN	Convolutional Neural Network
conv	Convolutional
EEG	Electroencephalogram
EMG	Electromyography
EPSPs	Excitatory Postsynaptic Potentials
ERP	Event-Related Potential
ERD	Event-Related Desynchronisation
ERS	Event Related Synchronisation
fMRI	Functional Magnetic Resonance Imaging
GP	General Practitioner
IASP	International Association for the Study of Pain
IPSPs	Inhibitory Postsynaptic Potentials
LSTM	Long Short-Term Memory
ML	Machine Learning
MRCP	Movement-Related Cortical Potential
MSK	Musculoskeletal
NPRS	Numeric Pain Rating Scale
PFC	Prefrontal Cortex
PNS	Peripheral Nervous System
RF	Random Forest
RNN	Recurrent Neural Network
SVM	Support Vector Machines

Contents

1 | Problem analysis

1.1 Introduction to Musculoskeletal Pain

This section gives a broad overview of Musculoskeletal Pain and its complications.
Musculoskeletal (MSK) pain refers to a broad range of conditions characterised by impairments in muscles, joints, bones, or connective tissue. It is the second leading cause of disability worldwide, affecting approximately 1.71 billion people or about 20 % of the global population.[1, 2]
MSK pain can vary in severity, presented as either mild or severe and lasting for short or long durations. Short term MSK pain is also called Acute MSK pain, this can e.g. be caused by torn muscles, muscle spasms, bone fractures, or joint dislocations. These injuries can affect all age groups and are experienced by almost everyone once in a while. Long term MSK pain, also called Chronic MSK pain, is defined as pain lasting over three months [3]. Chronic MSK pain can, e.g. be caused by sprained muscles, spinal stenosis, or lateral epicondylitis. Chronic MSK pain affects about 10 % of the world's population, where the most common types being lower back pain and neck pain.[3] The prevalence and types of chronic MSK pain vary by age and sex. Knee pain is e.g. more common in the elderly, while overall, MSK pain is nearly twice as prevalent in women as in men.[3]
Chronic MSK pain impacts daily living by leading to lifelong functional limitations, reliance on medication, reduced participation in daily activities, and the likelihood of sick leave or disability pensions. As a consequence, it also reduces the overall quality of life and represents a major public health risk, leading to substantial costs for the healthcare systems.[3]
In a study of the US medical economy by Zhou et al. [4], it was revealed that the US allocated 134.5 billion USD to treat lower back pain and neck pain. Furthermore, in the European Union, 2 % of the Gross Domestic Product of production cost was estimated to be lost due to MSK disorders.[4]
Therefore, MSK pain not only affects individuals by reducing their quality of life, but also places a significant burden on society through increased healthcare costs and lost productivity costs.
This leads to the initial research question, which is:

Initial research question

How is musculoskeletal pain perceived individually for each person, and how can it be measured and treated?

1.1.1 The perception of pain

What is musculoskeletal pain, how it is experienced by patients, a further understanding is provided in the section.

The general description of pain is described by The International Association for the Study of Pain (IASP), IASP defines pain as "*An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage*" [5]. Where MSK pain is described as acute or chronic pain in bones, muscles, ligaments, tendons, or nerves.[3]

Pain is complex because it is an individual perception, as the perception of pain arises from a combination of emotional experience, sensory perception, and physical sensation. Pain can therefore be described as how the individual experiences the situation, which makes pain a subjective experience.

Musculoskeletal pain is typically divided into three subgroups: Nociceptive, Neuropathic, and Nociplastic pain, which differ in their underlying mechanisms and causes. Nociceptive pain is the most common type of pain, it originates from the activation of nociceptors which are specialized sensory neurons that detect potentially harmful stimuli such as musculoskeletal strain or thermal harm. Nociceptive pain serves as a protective function, alerting the body to potential or actual tissue damage. In the context of MSK pain, nociceptive pain is often associated with conditions such as osteoarthritis, lateral epicondylitis, or muscle strains. Nociceptive pain can be further categorized into; Somatic pain, which arises from muscles, joints, ligaments, or bones, and Visceral pain, which is pain arising from internal organs. Within nociceptive pain types, somatic pain is the most prevalent and the easiest to locate, while visceral pain is typically more diffuse and difficult to pinpoint.[3]

Neuropathic pain, on the other hand, results from damage or dysfunction within the nervous system itself. Unlike nociceptive pain, which occurs in response to external stimuli, neuropathic pain can persist due to abnormal nerve signalling. In MSK conditions, neuropathic pain can develop from nerve compression, such as in radiculopathy due to a herniated disc. Because it does not serve a protective function like nociceptive pain, managing neuropathic pain can be more complex.[6] Neuropathic pain is challenging to detect because nerves can also undergo structural or functional changes due to plasticity. When discussing nerve plasticity, various subgroups can be identified, but a key aspect is neuroplasticity, which is the ability of the brain or Central Nervous System (CNS) to adapt, by altering the structure or function. This adaptation could occur as a response to nerve damage, where unaffected nerves structurally change to compensate for lost functionality.[7]
Nociplastic pain, is a different kind of pain as it could be the sensation of nociceptive pain even when the nociceptors detect no damage. This is because nociplastic pain is an altered change to the sensory pathway, which can happen both in the Peripheral

Nervous System (PNS) and CNS where a functional change amplifies the electrical signal describing the pain. The kind of functional change can be described by adaptive or maladaptive plasticity. Adaptive plasticity would be in a beneficial way where it, e.g. could restore nerve functionality through structural changes. Maladaptive could be when a change is happening, causing the nerve to be oversensitive or amplifying the signal, which could lead to a pain sensation, even when the sensation is not pain. These changes can sometimes result in persistent or abnormal pain signals, contributing to chronic nociceptive pain, as small stimulations would be experienced as a pain sensation.[7–9]

While musculoskeletal pain includes nociceptive, neuropathic, and nociplastic pain, this project will focus on nociceptive and nociplastic pain. Both are linked to the nociceptors, either through direct activation due to tissue damage (nociceptive pain) or the altered sensory processing which is causing the sensation of nociceptive pain (nociplastic pain).

Subjective musculoskeletal pain perception

How is musculoskeletal pain a subjective sensation and how is musculoskeletal pain perceived, are two questions answered in the following section.
The subjective perception of MSK pain arises from the combined activity of multiple brain regions stimulated almost simultaneously as a response to a potential pain stimulus.

Consider the following example: A strained muscle in the hand triggers a sensory response. The nociceptors detect the acute nociceptive pain, generating an action potential which travels through the PNS via the first-order sensory neuron. This action potential enters the CNS by reaching the spinal cord through the dorsal root. Within the spinal cord, the action potential is transmitted to a second-order neuron, which begins in the grey matter. After decussating within the grey matter, the action potential ascends further through the white matter. The action potential then ascends within the white matter through the spinal cord and CNS to the thalamus, which processes and filters sensory inputs. The thalamus acts as the brain's sorting centre, the thalamus directs the action potential via a third-order neuron to the somatosensory cortex, where the precise location of the pain is mapped.[10] This pathway can be seen in Figure 1.1.

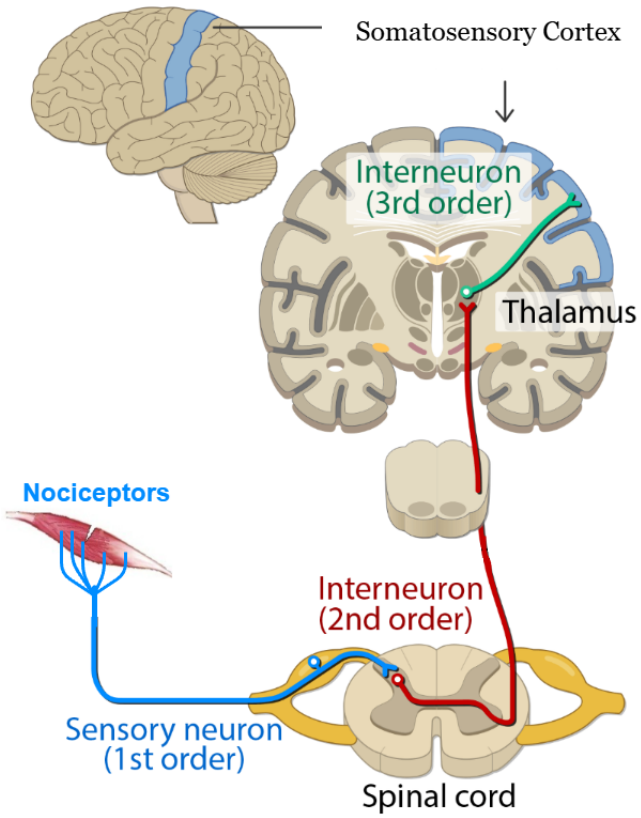

Figure 1.1: *Illustrates the pathway of the nociception pain. From the nocicepters detecting the pain and sending it through the sensory neurons to the somatosensory cortex. The image is an altered version from OpenStax [11], with modifications to brain area labels and the representation of the nociceptors.*

However, the somatosensory cortex is just one of many brain regions involved in pain perception. The nociceptive sensory signal is also sent to the hypothalamus, anterior cingulate cortex, and amygdala, which are also involved in pain perception, along with numerous other brain regions [10]. The brain regions are sketched in Figure 1.2.
In the hypothalamus the action potential triggers the fight-or-flight response, alerting the individual that potential harm is happening.
In the amygdala, the sensory input is linked to emotional processing, assigning an affective dimension to the pain. When receiving nociceptive pain, this emotional response is often unpleasant.
In the anterior cingulate cortex, the significance of the sensory sensation is associated with previous experiences of this kind of sensation.
The posterior insular cortex processes the intensity of pain and the response from the amygdala, thereby combining the sensory and emotional responses.[10]

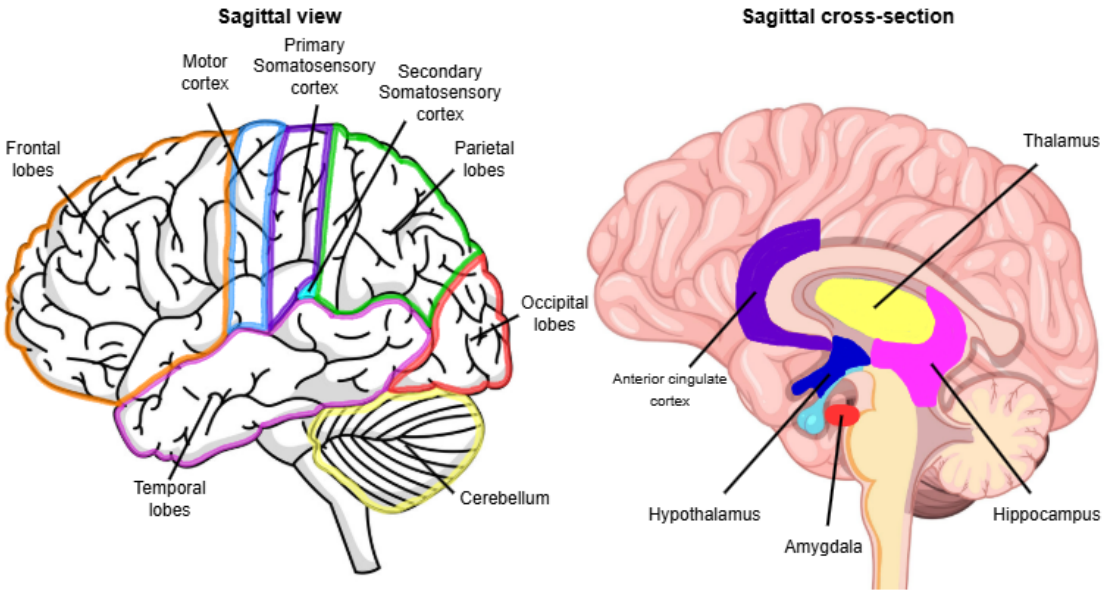

Figure 1.2: *Illustrates some of the brain regions included in pain perception. Each coloured and described with the name of the brain region.*

Due to the processing and transmission time of the action potential, an event-related potential (ERP) term describes the time from the noxious stimulus to the observable electroencephalogram (EEG) response in the brain. This ERP latency is approximately 250 ms for the subjective pain response.[12]

1.1.2 Diagnosing musculoskeletal pain

The challenges of diagnosing Chronic MSK pain due to neuroplasticity and the subjective perception of pain is described in this section.
The diagnosis of acute MSK pain is primarily assessed in the primary line of care at the general practitioner (GP). This diagnosis would be based on the duration, onset of the pain, the localisation, the pain intensity, and reviewing earlier medical records [13]. It is important to differentiate between pain and nociception. Pain is the combination of subjective feelings and the overall sensory experience, while nociception refers to the neural processes triggered by the tissue damage. To assess the intensity, duration, and onset of the nociception at the GP, a Numeric Pain Rating Scale (NPRS) is used, where the patient can describe the subjective pain from no pain at 0 to 10 which is the worst imaginable pain. The medical records are also reviewed as earlier diagnosed diseases could also be the cause of the underlying pain.[14] The initial diagnosis from the GP could lead to a referral to the more specialised secondary line of medicine, for further imaging or examination of the patient. The secondary line of care would include specialised professionals who can provide a more specific diagnosis or advanced imaging.

However, although NPRS might be a generally good way to diagnose acute MSK pain, the assessment and use of this application are not seen as reliable and accurate when diagnosing or estimating the cause of non-traumatic chronic MSK pain [15]. Furthermore, a problem which is seen is a diagnostic uncertainty which is generated at the GP and causes the chronic MSK pain patient to be referred back and fourth between the GP and secondary line of care for diagnostic examinations, without getting any further diagnostic information about the pain or how to manage it [13]. The challenge in diagnosing chronic MSK pain could arise due to nociplastic changes. Maladaptive plasticity is caused due to functional plasticity changes in the nociceptive pathways, which could cause a false chronic sensation of pain [8]. This suggests that maladaptive plasticity can distort the perception of the stimuli, amplifying the intensity beyond what is actually present. As a result, even mild stimuli can be perceived as nociceptive pain, a condition known as allodynia, which occurs primarily due to neuropathic nerve damage.[8] This phenomenon further complicates current diagnostic methods, as it makes it difficult to accurately determine the location and cause of the subjective pain, given the potential for maladaptive plasticity changes.[16]

There is currently extensive research focused on measuring nociception using a single objective measurement rather than relying on a subjective pain assessment method like NPRS. Instead, the goal is to have a technology that objectively measures the patient's pain. This could serve as a gold standard in clinical practice, not only for evaluating the effectiveness of medications but also for aiding in diagnosis and helping clinicians determine the best treatment options.[17, 18]

Currently, two promising technologies that objectively measures nociception are Functional Magnetic Resonance Imaging (fMRI) and EEG. Both methods measure the response to pain in the brain and are non-invasive, but the technologies are measuring by two different methods. fMRI has different settings which can be used, but the current standard for measuring nociception is the Blood-Oxygen-Level-Dependent (BOLD) image setting. The BOLD fMRI measures the hemodynamic changes in the brain after a stimulus. With that, the different areas of the brain which is activated during the pain sensation can be determined and illustrated. One of the advantages of this technology is the high spatial resolution as it can scan the entirety of the brain in one scan, and therefore eventually inspect deeper areas of the brain during the pain experience. Thus some of the limitations are the high cost and low portability, the MRI scanner is typically placed in a dedicated room to prevent external interference with the magnetic field and to ensure proper shielding. Another limitation is the image scanning process, as movement can cause noise artefacts in the images, and the chance of movement during pain is highly prevalent.[10]

The second method is EEG, which measures the electrical activity in the brain during the pain experience. Those electrical signals are generated by the electrical neurons activated, where patterns of synchronisation, desynchronisation or frequencies can be measured in specific places of the brain. One of the advantages of this technology

is the temporal resolution, as it can live measure for multiple hours. Furthermore, the measures can be taken in a variety of places without too much 50 Hz electrical noise, and the entire system has a low cost. However, one of the disadvantages is the spatial resolution as it only measures the electrical activity in the top layer of the brain towards the scalp, and therefore not the deeper brain area's electrical responses to nociception. Furthermore, each EEG electrode records electrical activity from a large area, meaning it can capture signals from multiple brain regions that are activated simultaneously.[10]

Thus fMRI has a higher spatial resolution, the information about the intensity of the pain and the placement is limited, as it only shows the areas of the brain which are using the oxygenated blood. In contrast, EEG can approximate the localisation of the pain sensation, with the electrical activity from the somatosensory cortex. Furthermore, the intensity of the pain can be approximated as the amount of electrical activity in different frequency bands.[10]
Furthermore, studies shows a promising development for EEG as a technology used in the case of rehabilitation, this rehabilitation could be e.g. for MSK pain in the form of neurofeedback, which is an EEG based system giving feedback based on the activity of the brain, with the purpose of alleviating the pain sensation in the long run [19, 20].

1.1.3 MSK pain treatments

This section provides an overview of the treatment options, including some pharmacological and non-pharmacological treatment options.

Treatment of MSK pain can be challenging as it is dependent on the pain onset, duration, and severity, which are all considered during the evaluation of the different treatment options.
Where it is often divided into two different methods either pharmacological or non-pharmacological. The pharmacological treatment method is for when the pain is so great that it prevents the individual from doing daily activities or therefore decreases the quality of life of the patient. Therefore, to alleviate the pain a variety of drugs can be prescribed depending on the severity, type of pain and the area affected. One common pharmacological treatment option for pain is morphine, which alleviates pain by reducing nociceptive signal transmission from the PNS to the CNS. This dampens the electrical activity associated with pain perception. However, while morphine is effective for acute MSK pain, it has shown limited effectiveness in treating chronic MSK pain.[3]
The non-pharmacological treatment methods focus on managing pain without drugs. E.g. these would be interventions such as physical therapy, acupuncture, and other interventions aimed at improving function, reducing pain perception, and enhancing the overall quality of life. The treatment option depends on the patient's condition, pain severity, and response to previous therapy interventions. E.g. Acupuncture is often recommended as a short term pain relief for chronic pain patients, whereas for

long term pain management, a common treatment option is therapeutic exercises to regain muscle strength and mobility.[3]

In many cases, treatment primarily involves pharmacological pain relieving medications. Thus, for chronic MSK pain, a non-pharmacological approach is often preferred to have the best chance to rehabilitate and therefore optimally in the end not be dependent on pain-relieving medication. As the pain-relieving medication will not rehabilitate the damaged area, it only relieves the patient from the pain.[3] Therefore, treatment plans often include a mixture of both pharmacological and non-pharmacological treatments, as this approach helps the patient rehabilitate with minimal pain. Furthermore, the mixed treatment is also a widely accepted and evidence-based method that has been shown to provide the best analgesia in the long term.[21]

This leads to the development of the search question, which is going to be used for the *current solutions in the literature* section, where a structured literature search is conducted to find relevant literature. How the structured literature search was conducted will be described in the later method section, which will include the search protocol and the search string used.

Search question

How can Electroencephalography be used as an objective method to assess musculoskeletal pain, and what are the relevant EEG biomarkers for musculoskeletal pain?

1.2 Current solutions in the literature

This section is structured to analyse the literature to find current solutions by using a structured literature search, which will be described in the method part of the project.

Through the structured literature search, the problem with the current golden standard of detecting pain in the primary line of care, is seen as a limitation in capturing the pain due to the complex nature of pain, and could be affected by the physiological and psychological statuses of the patient.[19, 22–24] Furthermore, the self-reported pain intensity scores are seen to poorly correlate with the stimulus given.[19] The current studies therefore, look into EEG as a promising objective measurement of nociceptive pain, due to the high temporal resolution, low cost, broad availability, and ease of data collection.[19, 23] Another application for an objective pain measurement method would be by using it on patients with verbal limitations, such as patients undergoing general anaesthesia, patients with cognitive disorders, or young children.[23]

The general brain areas which are active during perception and processing are important to know and understand, as the timing and synchronisation of the different active brain regions could be important in pain processing.[19] Therefore, a list to

summarise the different brain regions linked to pain processing, mentioned in the articles, is provided.

- primary somatosensory cortices S1 [19, 24–26]
- secondary somatosensory cortices S2 [24–26]
- anterior cingulate cortex (ACC) [19, 24, 25]
- anterior insular cortex (AIC) [19, 25, 26]
- prefrontal cortex (PFC) [24, 25]
- insular cortex [19, 25, 26, 26]

The brain regions are each responsible for a specific part of the pain processing, where the different frequencies in the brain regions also have a role in the understanding of pain. As Rockholt et al. [19], Vijayakumar et al. [25], Martucci and Mackey [26] write, the S2 and anterior insular cortex, are regions correspondent to the level of the unpleasantness of pain, where regions like S1, ACC, and insular cortex are regions active during noxious stimulations, where it can be seen in the power of the theta and gamma frequency bands.

Activity in the anterior insula, frontal operculum and pons was related to pain intensity in chronic pain patients [27].
Thus, those are some of the regions active with a specific frequency, there remains a lack of clarity in the literature describing the areas of when synchronisation and desynchronisation occur and the frequencies active during the sensation of pain, whether it is acute, chronic, or experimental induced pain situations [24]. Acute and chronic pain do not necessarily share the same biomarkers, as their biomarkers are often distinct [24].
Where a broad range of EEG pain responses can be seen in the different articles, e.g. Rockholt et al. [19], Gervasio et al. [20] describes an elevated gamma and theta power as a pain response. Where Zhang et al. [24], Van Der Miesen et al. [28] describes an increase in beta power, with a decrease in the delta, theta, and alpha power. Furthermore, Segning et al. [29] describes a statistical difference seen as an increase in the alpha and beta bands.
In other articles, the pain perception is linked to alpha-band Event-Related Desynchronization (ERD), but the alpha band also gets activated during non-painful sensory stimuli [24, 30]. Whereas, a gamma band Event Related Synchronisation (ERS) is more specific to pain and correlates with the perceived pain intensity [24, 28, 30].

Therefore, pain perception is often reflected in various frequency bands, though there is disagreement in the literature about which ones are most relevant and also if an increase or decrease is the correct way to describe it. However, most articles describe multiple frequency bands contributing to pain classification, with variations depending on the stimulus type [24, 28, 31].

In chronic pain, a frequency change could be seen as a decrease in the alpha power [19, 30], where Seok et al. [32] describes it as an increase in theta and alpha bands with a decrease in the amplitude of the evoked potentials.
Where, during acute or experimentally induced pain, a change in the frequencies was observed by; a suppression in the alpha band oscillations [24], and 52 % of the studies in Mathew et al. [30] showed a decrease in alpha power, where 28 % of the studies in Mathew et al. [30] showed an increase in the alpha power. Furthermore, Mathew et al. [30] found that high gamma power is an indicator of perceiving electrical stimulation as a response to nociceptive pain.
Another thing to take into account is the pain evoked event related potentials, these can be seen as an increase in the power of the theta band [19] and high gamma band [19], and as a result of movement evoked pain an decrease can be seen in the gamma band [30]. However, the frequency differences are not the only information the event-related potentials can give, as the N2 and P2 waves are well documented to be associated with pain perception of high and low pain states seen in the supplementary motor cortex [24].
For the electrode placements of the different articles, a general use of the 10-10 electrode placement method was used [20, 29, 32, 33]. The article by Seok et al. [32] measured with a total of 24 electrodes, with 7 electrodes in the frontal, 4 electrodes at the somatosensory, 3 central, 4 temporal, 3 parietal, and 3 occipital. In Segning et al. [33], 4 electrodes were of interest, which were placed in FC6, T8, FC5, and T7. And lastly, Gervasio et al. [20] and Segning et al. [29], electrodes of relevance were those above the primary somatotopic brain region, whereas in Segning et al. [29] with a right-handed population it was electrode FC5 which was kept for analysis.

The reviewed articles performed preprocessing before analysing the data. This preprocessing included applying frequency filters to remove noise artefacts and selecting the time window for analysis. The choice of frequency bands varied between studies, leading to differences in filter cutoff points. E.g. Vijayakumar et al. [25], Miller et al. [31] applied a bandpass filter ranging from 1 to 80 Hz, while studies examining a broader frequency spectrum used a 1 to 100 Hz bandpass filter [25, 32]. In contrast, Vijayakumar et al. [25] applied a narrower bandpass filter from 0.16 to 43 Hz.

The methods for selecting the frequencies also differed across studies. An article employed Fast Fourier Transform [29], whereas Vijayakumar et al. [25] used the Gabor wavelet package with 60 different wavelets, covering a frequency range of 2 to 80 Hz.

These frequencies were analysed in Gervasio et al. [20] in some specific time windows, to analyse if event-related pain responses could be seen in the signal. This was done by segmenting the EEG signal into 6 second windows, 4 seconds before and 2 seconds after movement initiation.

This approach, with comparing and preparing the data, was seen both in studies manually looking at the data in the different frequency bands, but also in studies

using machine learning (ML) algorithms to extract EEG features, classify if pain is present or not, and predict pain intensity.[19, 22, 25, 34] One approach to prepare the EEG data for ML was to normalise it, as the EEG data can vary in amplitude with age. Therefore, a min-max normalisation was implemented in two articles.[29, 33] Various ML models were used in the articles, including classifiers based on support vector machines (SVM)[19, 22], Random forest (RF)[25], mini-ROCKET[34], and multivariate pattern analysis [25]. Additionally, Rockholt et al. [19] described the use of a state space model which is an unsupervised ML model, which requires fewer training trials with a comparable outcome to supervised ML models.

With the knowledge gained from the structured literature search, a further understanding of some areas is needed before the method section of the report can be structured.

1.3 Deeper understanding of EEG and how electrical signals are generated in the brain

This section provides a deeper understanding of what EEG records and what part of the brain generates this signal.

As outlined throughout the problem analysis, EEG records electrical activity from the outer layers of the brain. While this activity is ultimately triggered by nociceptive input, the EEG does not capture the original action potentials from nociceptors themselves, as these peripheral signals are too short.

Instead, the primary contributor to the recorded EEG signals is the action potentials from the cortical neurons within the cerebral cortex.[35] The cerebral cortex is the outermost layer of the brain and plays a crucial role in the communication between different brain regions. It consists of a network of interconnected neurons that transmit action potentials from one part of the brain to another. This is important for the brain regions in the cerebral cortex, such as the somatosensory and primary motor cortex. This makes it possible for these regions to relay and amplify signals, activating even more brain areas.[35, 36] At the core of this process are pyramidal neurons, which plays a crucial role in transmitting these signals and are responsible for the majority of the EEG signal.[36]

The pyramidal neurons have a triangular shape with the top of the dendrites pointing perpendicular to the scalp, and the base dendrites oriented deeper in the cortex. When a large group of pyramidal neurons is activated synchronously, a dipole is created, with the negative end near the scalp and the positive end deeper in the cortex.[36] The EEG captures the summed excitatory and inhibitory postsynaptic potentials (EPSPs and IPSPs) occurring at the dendrites of these neurons, demonstrating the overall synchronised neural activity. EPSPs depolarises the neuron and thereby increase the likelihood of an action potential being able to trigger neuronal firing. Conversely, IPSPs, hyperpolarise the neuron, making it less likely to fire.[37]

For EEG to detect activity, a large population of pyramidal neurons must receive synchronised EPSPs, generating a collective electrical dipole strong enough to be

captured by the EEG. If excitation and inhibition occur randomly, the signals cancel out, and no clear EEG activity can be seen. Thus, EEG primarily reflects the large-scale coordination of neural inputs rather than individual neuron firings, where the frequencies in the EEG signal are caused by synchronised activation of pyramidal neurons, with their EPSPs and IPSPs generating oscillatory patterns detectable by the EEG.[36, 37] The different oscillations are analysed by the frequency and as written in section 1.2, the different frequencies can represent different patterns, relevant to the detection of pain.

Another measurable feature that EEG can detect is something called ERP, which are electrical responses in the cerebral cortex triggered by sensory, cognitive, or motor events.[38] ERPs are produced by the pyramidal neurons firing synchronously while processing information and sending electrical signals between each other. One part of the ERP related to movement initiation is the motor movement preparation phase called movement-related cortical potential (MRCP), which is visible around at least 500 milliseconds before a voluntary movement [39]. MRCP is associated with the planning and execution of a movement and can be seen as a slow decrease in the amplitude of the EEG signal. As the motor cortex is as close to the somatosensory cortex, these electrical activations of the cerebral cortex can be obtained in the electrodes over the somatosensory cortex.[39]

1.4 Problem summary

To sum up, MSK pain is the second leading cause of disability worldwide, effecting about 20 % of the world population. Pain is generally divided into two groups: chronic pain and acute pain. Each of them with different treatments and treatment options. Thus, a general problem is the lack of an objective method for differentiating the types of pain, as current approaches are influenced by the psychological factors that may affect the patient's experience and reporting of pain. An objective approach to pain differentiation could therefore lead to a more accurate assessment, as it would minimise the influence of the individual's psychological state.

Thus, one question remains of how can pain be objectively differentiated, and what biomarkers should be looked for when detecting pain. The question is a widely discussed topic with different studies providing evidence of different frequencies or placements for features of pain detection. Multiple ML approaches have been applied to both classify and differentiate the pain types, thus without an approach as the best or without a clear answer.
The aim of this system is therefore to classify chronic and acute pain, with the purpose of differentiating them and finding the placements of where pain related features are present.
This leads to the following research question.

Research question: *How can a system be developed to classify chronic pain, acute pain, and healthy participants, and how can the temporal focus during the analysis be visualised?*

2 | Methods

2.1 Initial literature search and structured literature search protocols

This section provides information about how the literature search has been structured and conducted for the problem analysis.

The purpose of the subsection 1.1.1, subsection 1.1.2, and subsection 1.1.3 was to gain a deeper understanding of the issue by exploring the initial research question, from section 1.1: *How is MSK pain perceived on an individual level, and how can it be measured and treated.* Specifically, areas like MSK pain, why MSK pain is subjective, the diagnostic methods, and treatment options for MSK pain, were researched through an unstructured literature search process.

Through the problem analysis, a deeper knowledge of the problem area was gained, which formed search terms in the form of specific interventions and issues. This leads to the search question, which was used for the *current solutions in the literature* section of the project, by making a structured literature search of the current articles describing the search question, from subsection 1.1.3.

Search question:
How can Electroencephalography be used as an objective method to assess musculoskeletal pain, and what are the relevant EEG biomarkers for musculoskeletal pain?

A structured literature search was made to find peer-reviewed articles describing the problem area and forming the initial technology description of the current solutions in the literature.

The structured literature search is done by using an altered SPICE framework, which is a framework based on the PICO model. The original SPICE model includes Setting, Perspective, Intervention, Comparison, and Evaluation. However, due to this project not having any technology to compare to, this part of the SPICE model is excluded. Furthermore, the Setting part of the spice model is also excluded as all settings are of interest for this project.

This concludes in the altered SPICE model being set up as table Table 2.1 illustrates.

Table 2.1: *Illustrates the altered SPICE framework for the structured literature search protocol, which was conducted in PubMed, Scopus and IEEE.*

Perspective (P)	Intervention (I)	Evaluation (E)
Musculoskeletal pain	EEG	Pain biomarker
Musculoskeletal pains	Electroencephalogram	Pain biomarkers
MSK pain	Quantitative electroencephalography	Brain-based biomarkers
Chronic pain		neurophysiological biomarkers
		Quantification

The SPICE model is set up in the same way as the PICO model, such that the boolean operator "OR" is inserted between the words within each column, whereas the "AND" boolean operator is inserted between the columns. The search terms identified using the SPICE model were extracted and combined to create a search string, which is then used to search through the PubMed, Embase, and IEEE sites. These sites were chosen as they together provide a wide variety of peer-reviewed articles describing both the medical and technological sites of the problem area. Furthermore, a new Primo Research Assistant from Aalborg Universitetsbiblotek, was used to find five articles, which were also included in the structured literature search process. This AI was used by inserting the search question into the search field, which then replied with 5 articles describing the Search question. This PRIMO generative AI was added to identify some articles which were not found by the rest of the sites.

The search string used on the three sites was the following:

("Musculoskeletal pain" OR "Musculoskeletal Pains" OR "MSK pain" OR "Chronic pain") AND ("EEG" OR "Electroencephalogram" OR "quantitative Electroencephalography") AND ("Pain biomarker" OR "Pain biomarkers" OR "Brain-based Biomarkers" OR "neurophysiological biomarkers" OR "Quantification ") NOT ("neuropathic")

The search string provided 25 Embase articles, 10 PubMed articles, on IEEE 2 journals and 1 conference paper were identified, and 5 articles were identified with the Primo Research Assistant.

The found articles were first sorted to exclude duplicates from the 4 different sources. Afterwards, the title and abstract of each article were screened to identify if the topic of the article was relevant to this project. Lastly, the full text was assessed, and some articles were excluded due to the full text not being available. The following Figure 2.1 illustrates the sorting process and the number of articles included in the literature review describing the current solutions in the literature.

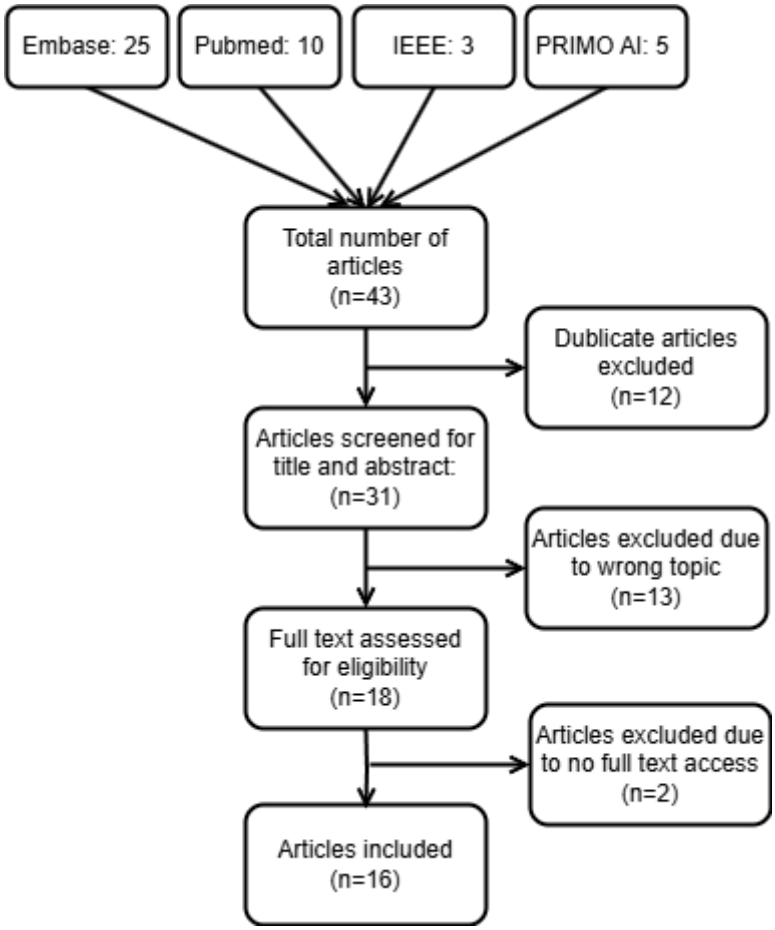

Figure 2.1: *Illustrates the literature sorting process of including and excluding articles.*

2.2 Data acquisition

This section provides information about the data which will be used for the ML part of the project.
The data consists of EEG and EMG measures from 32 participants, 15 of whom have chronic pain and 17 of whom are healthy. For each participant, a maximal voluntary contraction was measured as a base measurement, and afterwards, a recording of at least wrist extensions was made. The participants were instructed to sit with their dominant arm and elbow resting on a table. The movement involved lifting only the hand, thereby extending the wrist, while keeping the arm and elbow in contact with the table.
This movement was performed at least 20 times by each participant in the recording, who was instructed to have at least a seven seconds break between each movement. The movements were performed by all participants across the three groups: chronic pain, healthy controls, and acute pain.
The chronic pain group consists of participants diagnosed with chronic lateral epicondylitis, also known as tennis elbow. The healthy participants first completed the experiment without pain. Afterwards, they received a saline injection designed to temporarily replicate the pain associated with acute tennis elbow pain. These now acute pain participants then repeated the entire experiment under the same conditions, but while experiencing induced acute pain.

This overview of the different participants can be seen in Figure 2.2.

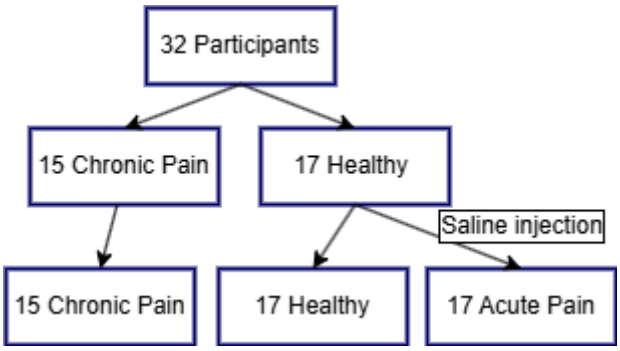

Figure 2.2: *Illustrates the three groups, and how the amount of data is divided. Furthermore, a line can be seen, describing when and to whom the saline injection was given to.*

2.3 Machine learning approach

This section gives an overview of the current ML approaches used for pain detection. Numerous ML approaches have been applied to EEG data analysis. Commonly used models include SVM, Multivariate Pattern Analysis, K-Nearest Neighbors, and Decision Trees.[28] Whereas, in pain related EEG studies, the two most commonly used ML paradigms are supervised and unsupervised learning.[19]

In a supervised ML approach, the model is trained on a dataset that includes labels, which are known outputs or characteristics describing each data point. These labels serve as references that guide the model during the training iterations, allowing it to learn patterns that can be associated with specific outcomes or classes. The goal is often to classify new data or predict values that resemble the provided labels. This approach, therefore, requires prior knowledge about the data to set up the labels needed.

In contrast, an unsupervised ML approach does not rely on labelled data. The model is trained using only the raw input data and attempts to identify underlying patterns, groupings, or features without any guidance in the form of labels. As a result, it is often used for exploratory analysis or feature extraction. However, besides this ML approach requires more advanced algorithms to identify patterns without prior knowledge, can it also offer a more unbiased analysis of the data.[40]

Given the varying strengths and limitations of the supervised and unsupervised ML models, selecting the appropriate ML model is essential.
E.g. an Unsupervised convolutional variational autoencoder is a complex ML model capable of analysing and replicating a signal. It can identify and extract the components that best represent the entire signal. This method produces a certain number of features which represent the input signal and reconstruct it only by using these variables. The features that are used for the reconstruction can be saved and used for a classifier afterwards. One disadvantage, however, is that the features are hard to interpret, as it can be challenging to identify where the feature is extracted from in the signal and what the feature values represent in the signal.[41]
Another ML example is the supervised convolutional neural network (CNN). This model analyses signals through training iterations and learns to extract features based on labelled data. This means that the outcome or category of each signal is known during the training, and the model then trains to identify features in the signal connected to the label for the signal. The CNN architecture consists of various layers, each with specific functions such as feature detection, downsampling, or determining how much the layer should impact each feature. While this approach often leads to highly accurate classification or prediction of results, it is influenced by the quality and consistency of the labelled data.[42]

Given the knowledge about the dataset with the three distinct classes labelled, the limited amount of data available in the dataset, with the knowledge about the two ML approaches, it is decided that the most suitable approach is a supervised CNN.

2.4 Programming of the system

This section provides information about the programming of the system developed to analyse the data, which includes data selection and preprocessing, but also programming of the CNN.
The system was programmed on a local laptop, while training and testing were performed using UCloud. The UCloud service provided access to four NVIDIA A10 GPUs, each with 24 GB of VRAM. With the use of the UCloud accessories, it accelerated the training and testing process by utilising GPU computation through PyTorch's CUDA support. The entire system was built and tested with Python version 3.12.7 and CUDA version 12.6.

2.4.1 Dataset

The dataset used in this study was recorded using the procedure described in Section 2.2. It consists of 49 recordings, each containing at least 20 movements, categorized into three classes. Recordings from the first 28 participants were collected using a single device, while the recordings from the last 4 participants were collected using two devices operating synchronously. The single device setup recorded data at a sampling rate of 256 Hz, which included 10 EEG channels and one EMG channel. In contrast, the dual device setup recorded at a sampling rate of 512 Hz and included 20 EEG channels and two EMG channels.
The EEG channels in the dual device setup included the same 10 channels as in the single device setup, with an additional 10 channels placed around them. The additional EMG channel in the dual setup was placed slightly more proximal than the first EMG channel, which was positioned in the same way as in the single device setup.

2.4.2 Channel selection and preprocessing

The selection of the EEG channels is important, as that is the data the CNN is trained with. The selected channels are Cz, C3, C4, Fz, and Pz, as they provide information about the electrical activity in pain specific regions of the brain. The Cz, C3, and C4 electrodes represent the activity in the somatosensory cortex, which are frequently used electrodes in pain studies, as the section 1.2 also describes. The Fz electrode was selected as it provides electrical information from the insula, anterior cingulate cortex, amygdala, and the prefrontal cortex. All areas are mentioned in the section 1.2 as potential biomarkers for pain, or that the area is active during the sensation or processing of pain.[43] The last electrode selected is the Pz, this is included to visualise the activity in the secondary somatosensory cortex.[43] The electrodes chosen can be seen in Figure 2.3.

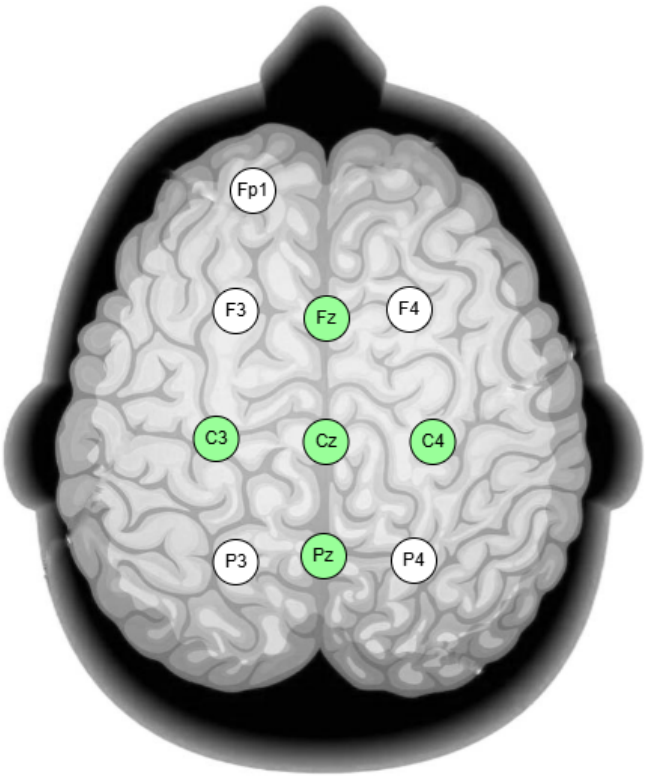

Figure 2.3: *Illustrates the EEG channels that the data was recorded with. Green channels indicate those used in the analysis, while white channels represent available but unused channels.*

Resampling, segmentation and unit conversion

The segmentation code will be described step-by-step, with some code examples.

To ensure the data was uniform, a resampling code was generated to ensure all the data was 256 Hz. The resampling code utilised the Scipy package's function *signal.resample*, which resamples the EMG and EEG data with the use of a Fourier method.

During the manual inspection, it was also observed that each movement recording from the participants was measured in one continuous recording, meaning the data should be segmented into smaller recordings, each segment including only one movement.
Furthermore, a filter was applied to the EMG signal for better analysis and visualisation. The filter used was a 4th order Butterworth highpass filter, with a cutoff frequency of 20 Hz, inspired by the article Gervasio et al. [20]. And for the sake of the analysis and making the data uniform, a baseline adjustment was also applied to the EMG signal, which takes the mean value of EMG signal down to a mean amplitude of 0. An illustration of the EMG data, with the applied filter and baseline adjustment, can be seen in Figure 2.4. The EMG signal includes some redundant data before the experiment, which also can be seen in Figure 2.4. This meant that it

was necessary to define a start index indicating when the segmentation code should begin identifying movements.

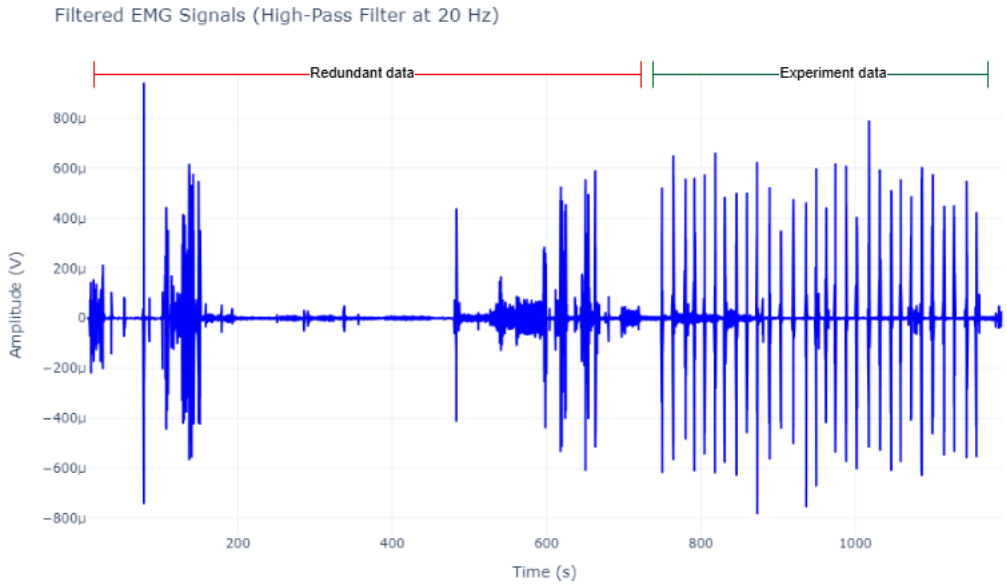

Figure 2.4: *Illustrates the EMG data, where the entire signal can be seen recorded in one continuous recording. This also includes some redundant data, illustrated with the red line, and the actual data from the experiment illustrated with the green line.*

The segmentation code should therefore only start once the start index has been defined, which was set manually for each participant, as the redundant data was not the same for all participants. Afterwards, each movement initiation will be identified and used as an index to determine when the data from the channels should be saved. Therefore, a script that analyses the EMG data, identifies movement initiations, and marks each initiation as an index, will be described.

This movement initiation detection code was inspired by the article Hodges and Bui [44], which describes different techniques for movement onset detection, and describes a method called *computer-based onset determination*. This method uses a sliding window to find the movement initiation. For each step through the analysis, the average mean value and standard deviation value of the samples in the sliding window are calculated. The window, then slides one sample forward, and the new sample is compared to the values from the previous sliding window. If the value of the new sample loaded is within 5 standard deviations of the average value in the sliding window, it is added to the sliding window as a sample, and the statistical calculations are updated. If the new sample is more than 5 standard deviations away from the average value in the sliding window, it is marked as a movement initiation. After a movement initiation is marked with an index, a pause of 5 seconds was implemented to ensure multiple indices were not detected during the same movement.

The size of the sliding window was set to 7 seconds or 1792 samples, and the movement initiation detection code applied to the EMG signal, can be seen in Figure 2.5.

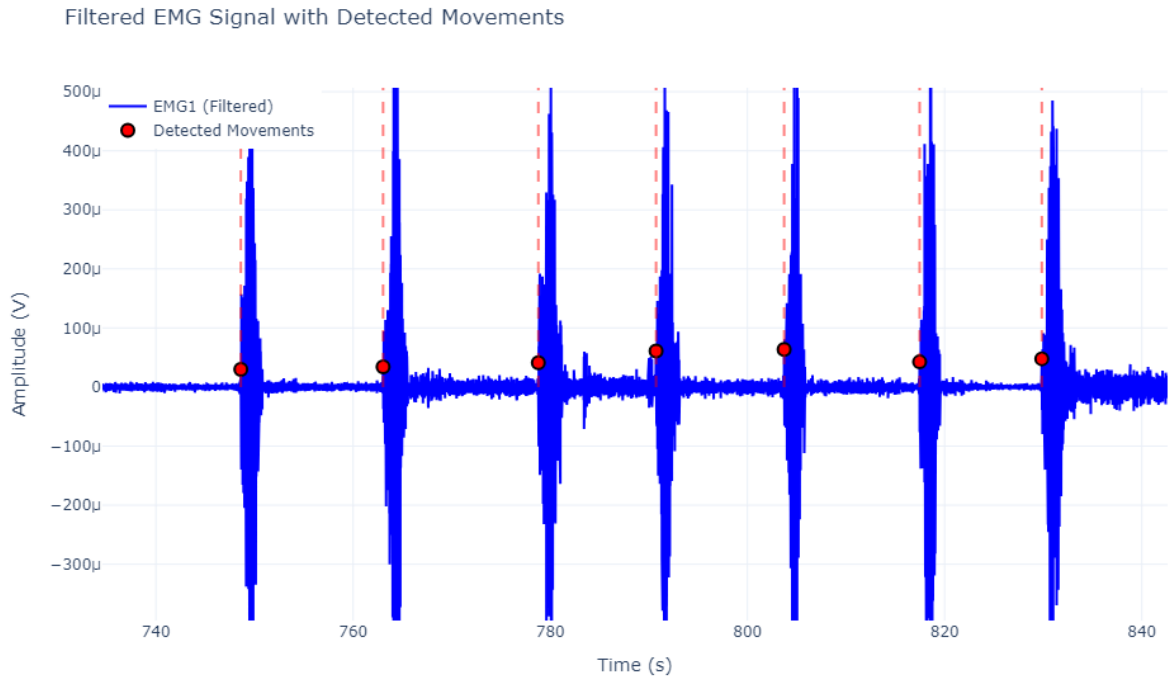

Figure 2.5: *Illustrates the movement initiation code applied to the EMG signal, where the EMG signal can be seen in the blue, and the red dots mark the movement initiations detected by the code.*

When the index of all movements has been marked, a selection process of 2 seconds before movement initiation and 3 seconds after movement initiation, thus a 5 second segment should be saved of the EEG data. This was possible since a temporal synchronisation was already done across the EEG and EMG devices. Therefore, the selected channels were saved.

Data sorting and splitting

After each segment was saved, a total of 1449 segments needed to be sorted into different groups to ensure the data was ready for training the ML model. This means the entire dataset of the segmented data was first split into 80 % training data and 20 % test data, and afterwards the 80 % of training data were split a further 80/20 into training and validation sets. The second splitting was made to have the validation set, which is used to evaluate the ML model during the training iterations and thereby generate a better ML model.

When all the segments of EEG data were split into different ML training groups, a code ensured that the data was in microvolts, adjusted to have a baseline of 0, and all segments were in a range of $\pm$ 150 μV. The data was converted to microvolts

by multiplying the channel's values by 10^6, as this would take the EEG channels' values from Volts to microvolts. The baseline adjustment was coded to ensure the data did not have a difference in the baselines of the channels, as EEG values from different participants could have different baseline values, due to gender and age.[29] The baseline adjustment was done by subtracting the mean value of the entire signal from each sample, and thereby adjusting the offset the signal might have, and if the signal did not have any offset this calculation would not do anything to the signal. The last part looked into excluding data where the values were outside $\pm$ 150 μV from the baseline. This was implemented to ensure data segments, as seen in Figure 2.6, were excluded.

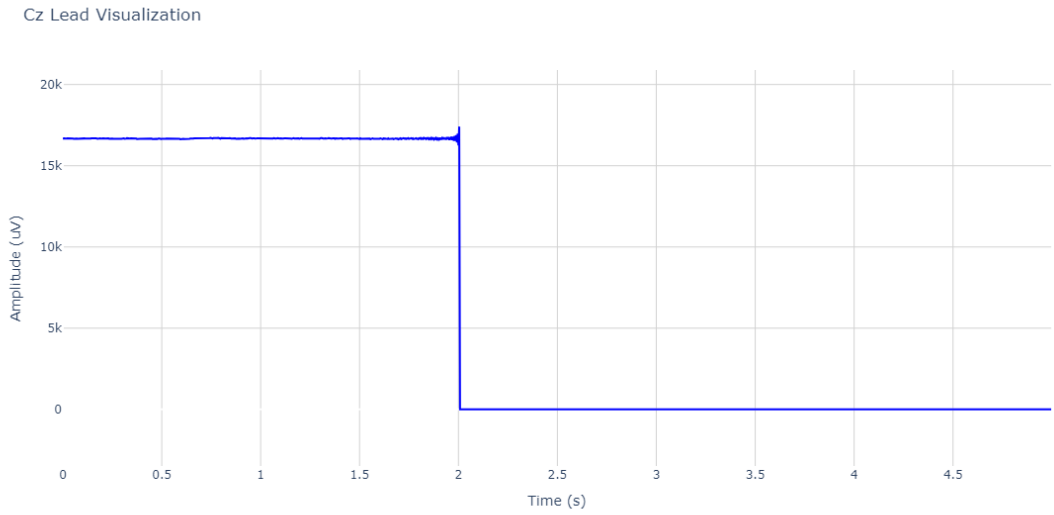

Figure 2.6: *Illustrates a data segment which was excluded due to the amplitude of the EEG channel.*

The values of the amplitude of the excluded data segments could be caused by electrical noise, muscle artefacts, or movement artefacts, resulting in unrealistic values and therefore unusable data segments.
The data exclusion resulted in a total of 66 segments being excluded.
Which means the final distribution of the data in the training, validation, and test sets can be seen in Figure 2.7.

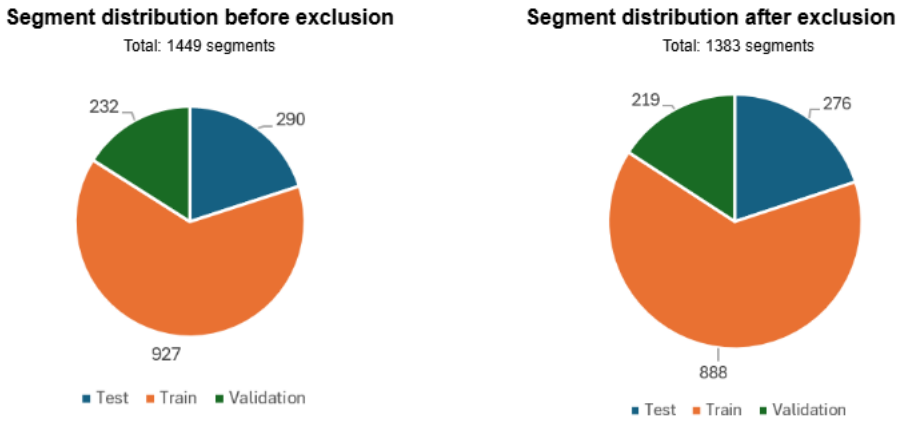

Figure 2.7: *Illustrates the number of data segments before and after the data exclusion.*

The data is then segmented, sorted, and split into datasets for the ML training. Therefore, the data of the total 1383 segments are ready to be used for the training and testing of the ML model.

2.4.3 Convolutional neural network

A CNN is a neural network and therefore first needs to be trained to analyse and find features in the data, by using the training and validation datasets. After the training is complete, the trained CNN can be used as a function where the test data is fed to it, and it replies with an answer in the form of which of the three classes it analyses the data to be in.
The architecture of the CNN is inspired by the article Yan et al. [45] and altered to include the functionality of a Long short-term memory (LSTM). LSTM models are a part of the Recurrent Neural Network (RNN), which are designed to process and analyse sequential data. RNNs can capture temporal patterns by passing information from one step to the next. However, standard RNNs struggle to retain long-term temporal patterns, due to functionalities like vanishing gradients. LSTM networks compensate for this limitation by implementing a memory cell which stores information over longer times. As a result, LSTMs can perform the same sequence analysis as RNNs while also effectively learning long-term patterns in the data.
The general setup of the CNN architecture is made on the same principles as Yan et al. [45], where the input signal is analysed in parallel by three different convolutions: small, medium, and large. And afterwards sent through the rest of the CNN, before getting analysed by the LSTM, in the end before classifying the data into a group.

LSTM CNN architecture

The general architecture and how the LSTM CNN is set up can be seen with the functions and a description of them, and the final architecture is displayed in the end of this section.

Convolutional layer

Convolutional (conv) layers are the backbone of the CNN, which are layers using a kernel of a certain size to analyse the signal. So a conv layer with a kernel size of 10, analyses 10 samples of the signal at a time and extracts important features from the samples. The extracted information could be frequencies or patterns in the signal. The code for the conv can be seen following code snippet:

```
1 Conv1d(channels, output_size, kernel_size, stride, padding)
```

The function Conv1d performs a one-dimensional convolution, and the inputs for the function are five different settings. The first input for the function is the *channels*, which are the number of channels from the EEG it should analyse, so in this case, five channels. The *output size* is the number of features it should learn from each conv performed, a larger number gives the conv a better chance to extract the perfect feature, but also makes more non-important features. The *kernel size* is the number of samples the conv should analyse at a time, so for a smaller conv it would be 5 samples, and for larger kernels it would be 21 samples. The *stride* input describes the number of samples the conv should jump before applying the next conv, so a stride of 1 would jump one sample before analysing the next samples. And lastly, the *padding* is the number of zeroes which should be added in the front and back of the signal. This is important as it ensures the first and last samples are also covered and analysed by the kernel. The amount of padding is often half of the kernel size, as it provides iterations for the kernel to analyse the beginning and end of the signal.

The conv layers for the LSTM CNN are set up to have three layers in the start, with different kernel sizes. This is made to analyse each EEG channel for frequencies and patterns in the signal of different sizes. A smaller kernel with a size of 5, is implemented for faster frequencies or local patterns in the channels. The medium kernel with a size 11 was implemented to find the broader patterns or a little lower frequencies than the small kernel. Lastly, the large kernel with a size 21 was implemented to find the broader patterns or slower frequencies in the EEG signal. This approach, by implementing mixed kernel sizes, was inspired by the article Yan et al. [45].

Batch normalisation

The function batch normalisation, or batch norm, is a function able to normalise the features from each layer. The batch norm, makes the training complete in fewer epochs, but takes a bit longer for the code to run, as it needs to normalise each time it is called. With the use of batch norm, there is a smaller chance for the training of the CNN to overfit the data. Furthermore, batch norm decreases the risk of an exploding gradient being developed in the many layers of the CNN. The batch norm

is implemented in the developed LSTM CNN after the initial three conv layers, and also after the conv layers which collect the features from the three initial conv layers.

Pooling layer

Another fundamental part of a CNN is the pooling layer, which reduces the number of features in the feature maps generated throughout the network. Pooling works by applying a kernel over the data and summarising the features within that region. In the case of max pooling, the maximum value within each kernel window is selected and fed further into the next part of the CNN. This helps reduce variability and makes the model more robust, but if pooling is applied too much, important features or subtle patterns may be lost.

In this project, max pooling is used in the LSTM CNN architecture. This means that only the maximum values from each pooling window are passed on to the next layers. The kernel size can vary, but to avoid discarding too much information, a kernel size of 2 is used.

Dropout layer

The dropout layer is what can give the CNN some variability in what features it focuses on. As the functionality of the dropout layer is not to let the different layers fully interact with each other, and thereby locking some nodes of each layer, to interact and receive information at certain times. The dropout function is given the percentage of nodes which should be locked each time, and the nodes are randomly selected to be locked. This functionality provides some variability to the CNN such that the same CNN architecture, trained on the same data, does not give the exact same output. Furthermore, the dropout layer makes the CNN have a smaller chance of overfitting the data.

Long short-term memory layers

The LSTM layers of the CNN are set up through the use of RNN layers. The RNN is a neural network type able to analyse temporal patterns on a deep neural network level.

After the CNN have extracted what it has seen as important features, those features are fed into the RNN layers. Then the RNN can analyse the features from the CNN and look at the temporal patterns in the signal and the features. Those temporal patterns and features are then processed and the weights of their importance are generated and fed into an attention mechanism. An attention mechanism is essentially comparing the different time points according to their importance. This creates a representation of what areas and features are the most important temporal patterns while filtering out less relevant information. Which essentially is the output of the RNN and what the entire algorithm is using to classify the data into the three groups.

Early stopping

This method was implemented such the training would not overfit on the data, meaning it would learn too much about the training data and therefore perform poorly on new unseen data. Therefore, early stopping was implemented, which can be simplified down to if the training and validation accuracy does not improve over a chosen set of epochs, then the training of the CNN will stop and the epoch where the

best training and validation accuracy was seen, is the one model which will be used. The number of epochs it should wait before stopping can vary, in this project, it was set to 10 epochs. Furthermore, the number of how much is seen as an improvement was set to 0.0001.

Fine tuning and optimiser

Fine tuning has been an ongoing part of developing the LSTM CNN model. This process involves adjusting various hyperparameters and components in the architecture to improve the model's ability to extract meaningful features from the data, without overfitting. One of the earliest adjustments made was to the optimiser. This adjustment was made to the Adaptive Moment Estimation or ADAM. The ADAM optimiser was chosen because of its ability to efficiently handle sparse gradients and adapt the learning rate during training. The ADAM optimiser combines the information from the gradient of the error and how much the error is changing from each iteration, thereby it can update the model's weights in a balanced way. One of the fine tuning adjustments done to the ADAM optimiser was the value of the learning rate. This was adjusted after the first models, as a too high learning rate would make the training unstable and therefore unusable, but also a too low one would make the model slow to converge. This learning rate was adjusted from 0.01 to 0.0001.

Furthermore, as a loss function for the model and the training, the function cross entropy was chosen, which is a function specially made for classification cases and fits perfectly with this multi-class classification case. The implementation of the cross entropy function helps the model provide better gradients when the predicted class vary from the true class, thereby improving learning efficiency and classification accuracy.

Heatmap

A heatmap was implemented to visualise which parts of the signal the CNN focuses on during the analysis. This was achieved with the use of the Torch and Plotly libraries, which allowed for the generation of the attention based visualisations. The concept of using a heatmap was inspired by the study conducted by Yan et al. [45].

The output consists of multiple figures, including examples from correctly and incorrectly classified segments. Each figure displays a heatmap, which represents the attention weights over time, which is computed by the RNN. In the heatmap, yellow regions indicates time segments which the RNN sees as important temporal places and therefore indicates a higher attention weight. While the darker blue regions is seen as less important and indicates a lower attention weight. Therefore, the heatmap highlights the areas of the model which is seen as important temporal features while darkening the temporal areas which are seen as less important to the RNN.

The output is divided into different figure types to improve understanding of the signal and the attention of the LSTM CNN. The first figure type is the solo segments,

which illustrates the input channels at the top in the subplot, and the heatmap from the specific segment analysis at the bottom. This first figure type also includes whether the final prediction of the segment was correct or incorrect.
The second figure type is for each class, where the heatmap from the average attention weights for that specific class is displayed. This is also divided into correct predictions, illustrated at the top of the figure, and incorrect predictions, illustrated at the bottom of the figure.
The third type of figure illustrates the average attention weights for the entire test set, with the correct and incorrect predictions divided again in a subplot.
Lastly, to understand the temporal difference between the attention weights for correct and incorrect predictions, a figure illustrating the difference in the attention weights over time can be seen. This is done by subtracting the incorrect average predicted attention weights from the correct predicted attention weights, which is done for each instance of time.

Confusion matrix
To better evaluate the performance of the developed LSTM CNN model, a confusion matrix is used and made as an output at the end of the code.
The confusion matrix provides a breakdown of the model's performance by comparing predicted classes to the true classes for each segment, showing how many predictions were correct and where the model predicted incorrect. This allows for the identification of specific misclassification patterns and is useful in a multi-class classification task, which this dataset is. The confusion matrix helps to assess how well the model distinguishes between different classes.

2.4.4 Summary of the CNN architecture

The LSTM CNN architecture can be seen in figure Figure 2.8, where the different steps from the input signal to the classification can be seen, with a slightly simplified twist to it.

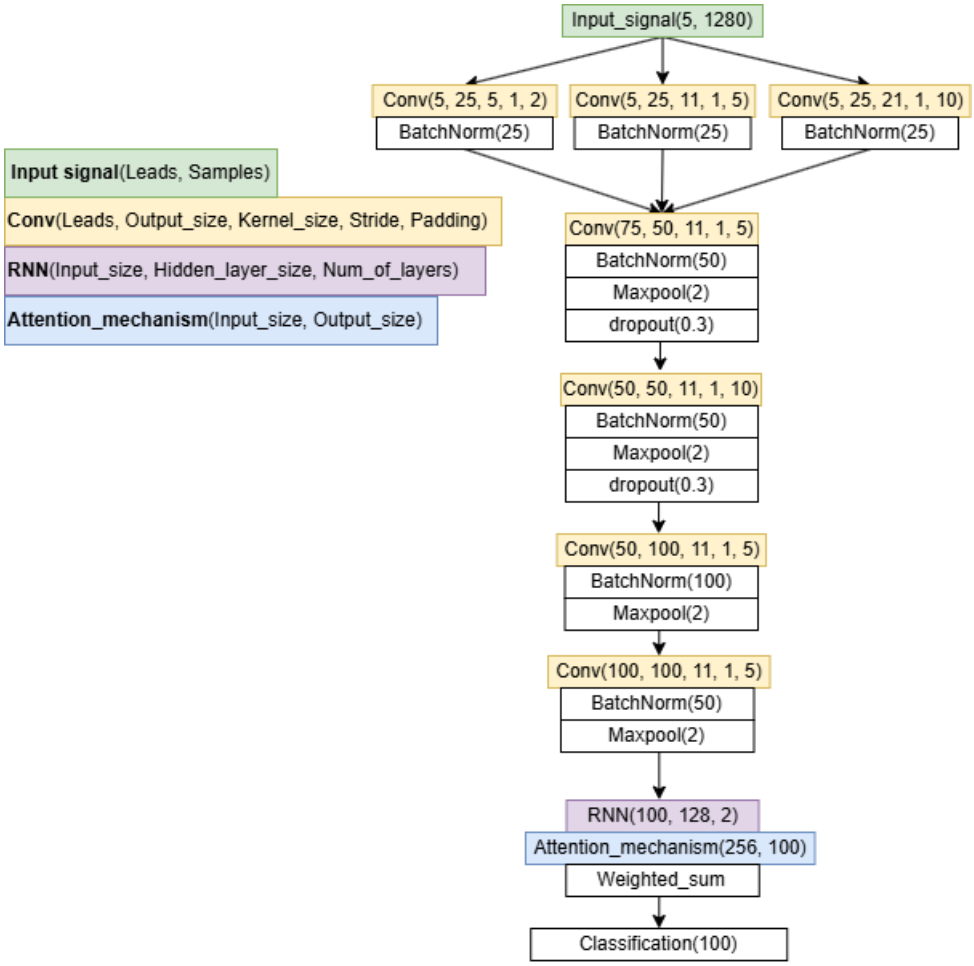

Figure 2.8: *Illustrates the architecture of the LSTM CNN model. The input signal is processed through three parallel convolutions, each using a different kernel size to capture different features. These branches are then merged and passed through a series of convolutions, normalisation, and pooling layers. The output is fed into the RNN layer, followed by an attention mechanism, and finally classified into one of the three classes.*

2.5 Testing the system

This section provides an overview of how the developed system was tested.
The testing of the system was performed on the, until now, unseen test set with the performance of the LSTM CNN as the result. During the training of the model, the same model architecture was trained five times, as the LSTM CNN includes some variability even between the same model trained on the same data, due to a combination of the dropout layers and the LSTM part setup through the RNN layers.
Therefore, five models were saved, and the model with the best validation accuracy will be chosen as the model to perform the testing on the test set.
After the training of the LSTM CNN is done, the weights from the layers in the

best epoch are saved. Afterwards the performance test loads the weights from the best epoch and uses them to make a static model of the LSTM CNN, which is then used for the classification on the test set. The predicted classification of the test set is then checked against the ground truth of the class, which will be illustrated in the confusion matrix. This is done to illustrate how well the model is at classifying the data and also to extract the prediction accuracy of the classification performed.

The output of the system is a combination of the accuracy numbers for the classes, heatmap images of the different types of figures described, a confusion matrix, a figure describing the training history of the accuracy and loss, and the general accuracy of the LSTM CNN on the test set.

3 | Results

This section provides the results of the test of the system, which is described throughout the method section.
The specifications of the five ML models trained are displayed in Table 3.1. The best model is selected and the output from it is illustrated beneath with the specifications.

Table 3.1: *Illustrates the training history for each of the models trained with the LSTM CNN and the test set.*

Model	Epoch	Train Accuracy	Val Accuracy	Train Loss	Val Loss
1	11	52.7	55.25	0.88	0.85
2	67	52.27	57.08	0.65	0.75
3	40	66.67	62.1	0.69	0.76
4	19	58.78	55.25	0.8	0.83
5	61	68.24	62.1	0.59	0.7

As seen in Table 3.1, model 3 and 5 both have the same validation accuracy of 62.1 %, which would be the specification that determines the model used for the results. Rather than choosing solely based on the validation accuracy, it can also be seen that the training accuracy is higher in model 5 than in model 3. Furthermore, it can also be seen that the validation loss from model 5 is the lowest of all the models trained. Therefore, model 5 will be the one chosen for the results section.

3.1 Accuracy- and loss model

The accuracy of the train and validation set during the training of the model, can be seen plotted with each other to the left in Figure 3.1. Where the training and validation loss can be seen to the right in the same figure.

Figure 3.1: *Illustrates the training history from model 5, which reached the best training score out of the five models trained.*

The best epoch, on which the model was built, was with the use of the weights from epoch 61. Where the training and validation accuracy from the best epoch, can be seen in Table 3.1.

3.2 Performance on test set

When the model was built and given the test set, the performance had an accuracy of determining the correct class on 54.55 %. This was with 150 correct predictions and 126 incorrect predictions of the class. The accuracy for each class can be seen in Table 3.2.

Table 3.2: *Illustrates the class accuracy when performing on the test set. The class can be seen to the left, including the accuracy of the class to the right. Furthermore, a parentheses with the correct predicted number of segments out of the total number of segments can be seen after the accuracy.*

Class	Accuracy (Correct/Total)
Chronic pain	75.68 % (56/74)
Healthy	13.76 % (15/109)
Acute pain	84.94 % (79/93)

3.3 Heatmaps

Heatmaps were displayed after the model had performed the classification on the test set and showed what the focus where on, when predicting the classes. The output of the LSTM CNN is a mixture of single segment predictions, class specific segments, and what the entire analysis focused on. The heatmaps of the single segments with the EEG illustrated in the top can be seen in Appendix A, the class predictions can be seen below with the average attention weights for the entire analysis by the

LSTM CNN.
The predictions of the three classes, both during the correct predictions and the incorrect predictions, can be seen in Figure 3.2, Figure 3.3, and Figure 3.4.

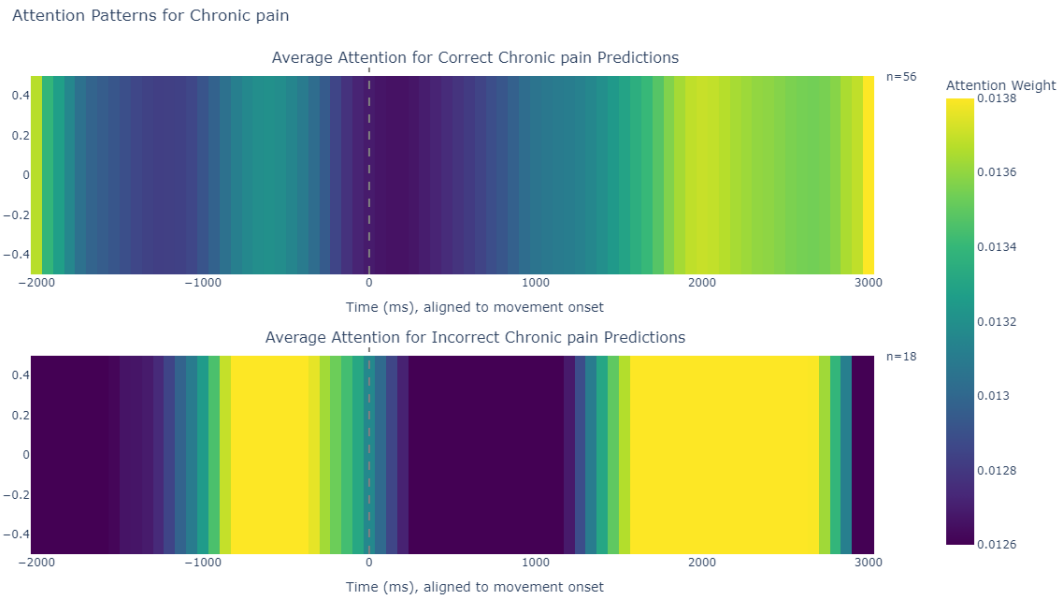

Figure 3.2: *Illustrates attention weights for correctly and incorrectly predicted chronic pain segments. The top panel shows the average attention for correct predictions, while the bottom part shows attention for incorrect predictions. Both plots are aligned to movement onset, with the number of segments (n) in each part's top right corner.*

What can be seen in Figure 3.2 is that the attention weights of the correct predictions of the chronic class are primarily focused on 4 intervals, from -2000 ms to -1933 ms, -666 ms to -466 ms, from 1800 ms to 2200 ms, and from 2866 ms to 3000 ms. The correct prediction attention weights are made up of 56 different data segments, whereas the incorrect predictions are made up of 18. The incorrect is showing two clear places of attention, from -866 ms to -266 ms and 1533 ms to 2800 ms.

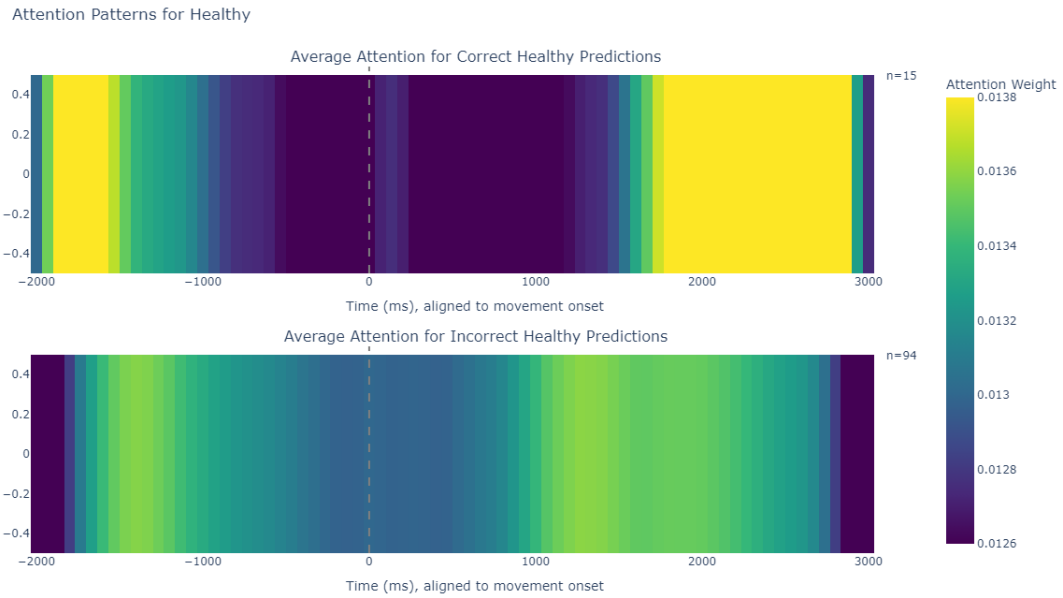

Figure 3.3: *Illustrates attention weights for correctly and incorrectly predicted healthy segments. The top panel shows the average attention weights from the correct predictions, while the bottom part shows the attention weights from the incorrect predictions. Both plots are aligned to movement onset, with the number of segments (n) in each part's top right corner.*

The attention weights of the predictions of the healthy class can be seen for the correct predictions, focusing on two spots, -1866 ms to -1533 and 1733 ms to 2933 ms. The correct prediction is made up of 15 data segments, and the incorrect prediction is made up of 94 data segments. The incorrect predictions are looking at almost the entire signal, thus with 2 spots more than the rest, from -1533 ms to -1200 ms, and 1066 ms to 1600 ms.

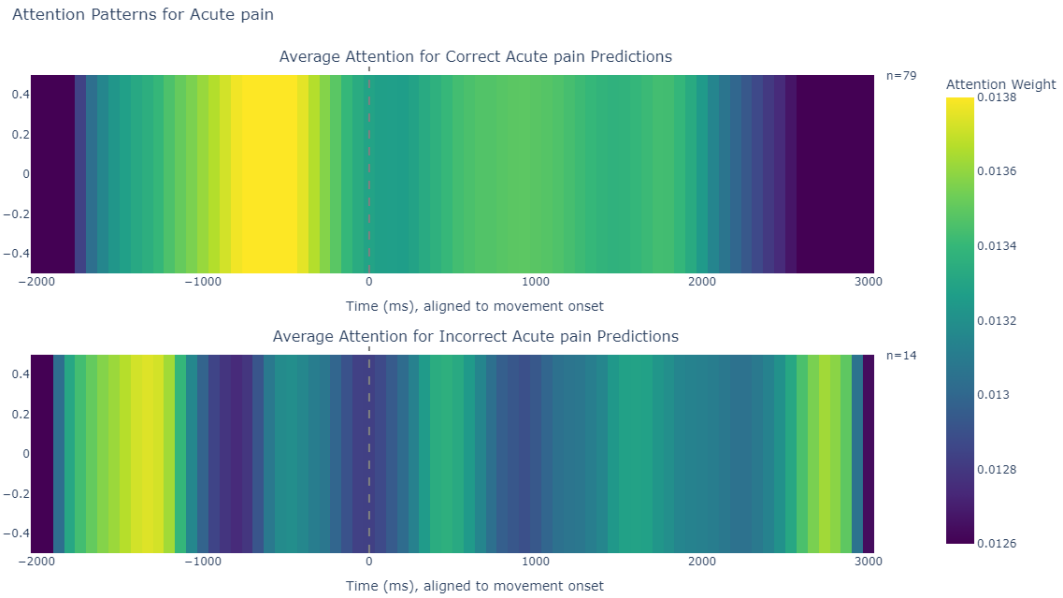

Figure 3.4: *Illustrates attention weights for correctly and incorrectly predicted acute pain segments. The top panel shows the average attention for correct predictions, while the bottom part shows attention for incorrect predictions. Both plots are aligned to movement onset, with the number of segments (n) in each part's top right corner.*

The attention weights from the acute pain participants can be seen for the correct predictions to be made up of 79 segments. The correct prediction heatmap illustrates one particular spot, which is highlighted from -800 ms to -333 ms. Unlike the incorrect predictions made up of 14 segments, which focus primarily on two out of the five spots highlighted in the signal, from -1600 ms to -1133 ms, and 2666 ms to 2933 ms.

The heatmap displaying the average attention plot for the entire LSTM CNN during the analysis of the test set, can be seen in Figure 3.5.

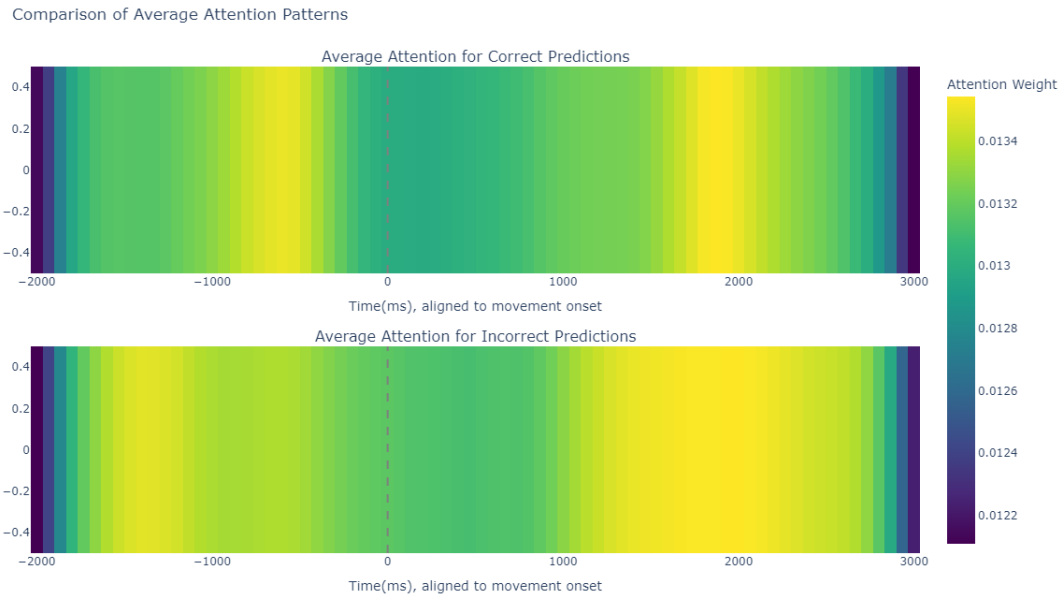

Figure 3.5: *Illustrates a heatmap of the CNN during the entire analysis of the dataset. With the correct predictions at the top and the incorrect predictions at the bottom part. The x-axis shows the time of the segment aligned to the movement onset.*

The attention weights for the correct predictions can be seen focusing on 2 specific spots, particularly, but overall with some attention to almost all parts of the segment, the 2 specific spots are from -800 ms to -466 ms, and 1733 ms to 2066 ms. Where the attention weights for the incorrect predictions are also focusing on two spots, but a bit broader spots, the first spot is from -1533 ms to -1133 ms, and the second is 1266 ms to 2400 ms.

Then the difference in the attention weights, for the correct and incorrect predictions, which can be seen in Figure 3.6.

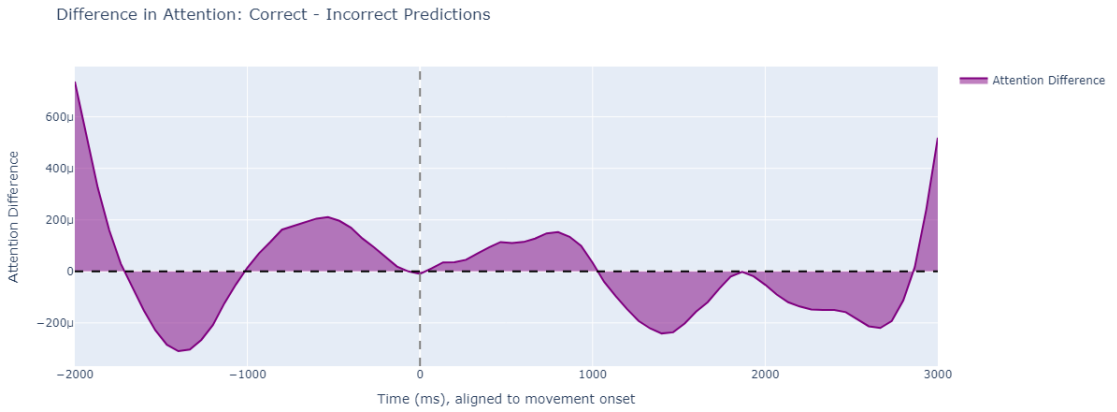

Figure 3.6: *Illustrates the correct and incorrect predictions subtracted from each other to visualise the difference in the attention weights over time for the entire LSTM CNN.*

3.4 Performance metrices

Lastly, as a performance metric of the LSTM CNN, a confusion matrix can be seen in Figure 3.7 for the predicted classifications and the true classes.

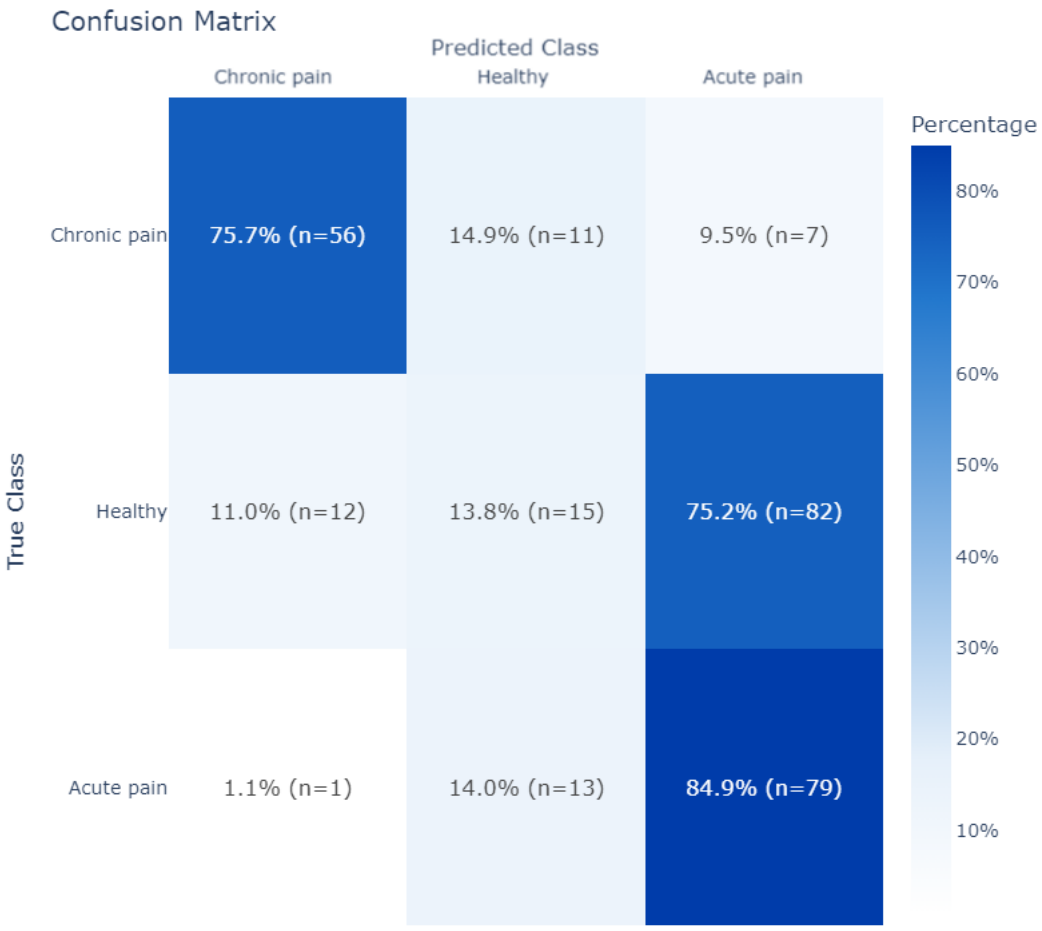

Figure 3.7: *Illustrates a confusion matrix with the percentages and the number of segments in each matrix part. The x-axis is the class predictions made by the LSTM CNN, and the y-axis is the true class for each segment.*

4 | Discussion

The study was structured to investigate the research question of: *How can a system be developed to classify chronic pain, acute pain, and healthy participants, and how can the temporal focus during the analysis be visualised?*.
With the aim of developing a ML model, able to distinguish the different classes and illustrate the temporal focus from where the system could differentiate each class. This was done with the use of the LSTM CNN ML model, analysing the dataset of 1383 segments, which resulted in an average classification prediction accuracy of 54.55 %. The used LSTM CNN model was the fifth model trained, achieved a training accuracy of 68.24 %, a validation accuracy of 62.1 %, a training loss of 0.59, and a validation loss of 0.7, as seen in Table 3.1.

4.1 Classification performance

The average of the classification predictions was also done on each class to find the class specific accuracies. Which resulted in a prediction accuracy from each class to be: chronic pain accuracy of 75.68 %, healthy accuracy of 13.76 %, and acute pain accuracy of 84.95 %, as seen in Table 3.2. This difference in class accuracy illustrates a limitation, as the healthy class weighs down the overall accuracy of the system.

Several studies have used ML models for pain detection, not only to classify pain presence but also to explore where in the data the pain related patterns are present. For instance, the systematic review by Rockholt et al. [19] examined SVM based models using different EEG frequency bands, noting that gamma and theta frequency bands were especially relevant. While initial models achieved an accuracy of 57 % with the use of two frequency bands, a later study combining multiple frequency bands, and thereby improved the accuracy up to 93.7 %, suggesting that pain related information may be distributed across the frequency spectrum, rather than localised in a single frequency.
In a study done by Tsigarides et al. [34], with a similar data size to this study, they used the mini-ROCKET ML model to classify healthy and chronic pain participants, with an accuracy of 70.7 %. This finding, from [34], suggests that both the size of the dataset and the preprocessing steps may play a role in increasing the predictive accuracy of the model. But taking into account the spectrum of the frequency bands, from the study Rockholt et al. [19], could be an important factor. The preprocessing part of filtering the frequencies, without cutting any important frequencies, could be one part of the potential future improvements for this LSTM CNN. Thereby, increasing the LSTM model's prediction accuracy.

4.2 Confusion matrix understanding and analysis

As seen in Figure 3.7, the classifier demonstrates a strong ability to detect the classes chronic pain and acute pain, with accuracies of 75.7 % and 84.9 %. In contrast, the healthy class is frequently misclassified, most notably as acute pain, with true positive predictions only made in 13.8 % of the cases. This suggests that the model struggles to differentiate healthy from acute pain, which could point to overlapping EEG features in the classes.

This misclassification could partly be explained by the fact that the healthy and acute pain segments were recorded from the same participants, with the only difference being a saline injection to induce acute pain. As a result, their EEG baseline patterns could share similarities, making it difficult for the model to differentiate the two classes.

With the numbers from the confusion matrix, the class specific Recall and Precision values can be calculated. The different calculations and formulas for Recall and Precision can be seen in Appendix B, where the calculation and formula for the F1 score can also be found.

The average Precision and Recall scores were both calculated to 0.543. Consequently, the F1 score, which reflects the values of the mean Precision and Recall, was calculated to 0.543. This F1 score can therefore be categorised as ok. Given the relatively low average Precision and Recall values, a high F1 score was not expected. However, an F1 score of 0.543 indicates that the model is not guessing randomly, as in a three-class classification problem, a model making random predictions would typically have an expected F1 score around 0.33.
Therefore, the model appears to be learning meaningful patterns from the EEG and making some correct predictions. Nevertheless, the performance still leaves room for improvement, particularly in trying to increase both the Precision and Recall values, and particularly the predictive accuracy of the healthy class. With such improvement, an increase in both the Recall and Precision would be expected, and therefore also an increase in the overall F1 score.

4.3 Class specific heatmaps

To further understand the temporal focus and places of the attention weights from the LSTM CNN, the heatmaps seen in Figure 3.2, Figure 3.3, and Figure 3.4 were analysed.
The first heatmap is from the chronic pain class, which can be seen in Figure 3.2. The focus of the LSTM CNN demonstrated through the attention weights, can be seen at the end of the signal from approximately 2933 ms to 3000 ms, where the attention weights are higher and therefore marked with a bright yellow colour, indicating a high

value of the attention weights. This placement of an important feature is unexpected when compared to the findings from Gervasio et al. [20], Sur and Sinha [38], Shakeel et al. [39], where the important features are seen from 600 ms before movement initiation to 1000 ms after movement initiation. Thus, the area around 2000 ms is also seen as important in the heatmap, which is at a lower attention weight, but is closer to the expected, than the highest value of the attention weights.
Compared to the average attention weights of the incorrect predictions in the chronic class, a clear attention towards 2000 ms ±500 ms, including the area around -600 ms. These predictions are much clearer in the areas of interest, but could be because the incorrect plot is an average of 18 segments, where the correct predictions are made up of 56 different segments, which can also be seen with the attention weights being a bit more vague and spread out.

In the attention weights from the healthy class, from Figure 3.3, a clear area of interest can be seen in the correct prediction, which again could be explained by the plot being made up of a total of 15 segments. But, when comparing the attention placement of the correct prediction from the healthy class to the incorrect predictions from the chronic class, a similarity of focus can be seen around 2200 ms. This similarity could be caused by 11 of the chronic pain segments being incorrectly predicted as healthy. As the placement of the attention weights in the incorrect chronic heatmap matches the placement of the attention weights of the correct healthy heatmap, and would therefore incorrectly be predicted as healthy segments. Another similarity between two classes' attention weights, can be seen in the incorrect healthy and correct acute pain. A similarity in the attention weights of two areas around -1533 ms and 1066 ms, which could be the cause of the many incorrect predictions of the healthy class. The similarity and incorrect predictions could also be because the acute pain participants and the healthy participants are the same. The only difference is the saline injection before the acute pain recording. Which could mean these two classes have too much in similarity, regarding their baseline EEG characteristics, to differentiate the healthy from the acute pain. Another explanation could be that during the recording of the healthy class, the participants experienced acute pain, due to muscle fatigue as a result of the repetitive movements and therefore actually ended up with acute pain EEG signs, making the differentiation of the two classes difficult for the LSTM CNN.

What generally can be said about the heatmaps is that the average attention weights are more vague and spread out when more segments contribute to the heatmap. This pattern can be seen in the correct chronic pain, incorrect healthy, and correct acute pain heatmaps. Where multiple parts of the heatmap consist of attention weights that are halfway up the attention weights scale. In contrast to the heatmaps with fewer segments contributing to them, as seen in the correct healthy and incorrect chronic heatmaps, where a clear areas of interest in the attention weights can be seen.

Overall, this suggests that the LSTM CNN prioritises the placement or pattern of

the attention weights, serving as an important differentiation factor between classes. However, there is still an uncertainty regarding how the different segments are being analysed differently, and therefore gives attention weights in unexpected and incorrect places. As seen in the incorrect chronic heatmap, the pattern of the attention weights is the same as in the correct healthy pain heatmap.

Therefore, a potential direction of future exploration could be to investigate the different segments that make up the different heatmaps. If there is a trend showing, this could e.g. be if the incorrect predictions of the chronic pain class were all segments from early in the recording, illustrating that the chronic pain occurs as a result of repetitive movements. This would give an insight into how consistent the data is in showing class specific patterns.
Furthermore, a potential exploration could be by having a new recording of the acute pain class, to see if the basic EEG characteristics had an influence on the current trained model, and therefore would make the model better at differentiating the acute and healthy classes.

4.4 Placement of the attention weights

Another thing the heatmaps illustrate is the main focus through the entire analysis of the segments, as seen in Figure 3.5. The main focus during the correct predictions of the segments is seen at the top of the figure, with two temporal spots with higher attention weights. The two spots are around -666 ms to -466 before movement onset, and 1733 ms to 2066 ms after movement onset.

As mentioned before, in the studies Gervasio et al. [20], Sur and Sinha [38], Shakeel et al. [39], the important features for pain related movements are seen to be from 600 ms before movement onset to 1000 ms after movement onset. The study [39], describes the area of movement preparation before movement onset, and the studies Gervasio et al. [20], Sur and Sinha [38] describes the pain related processing done by the different brain areas, as the spot after movement initiation.

When comparing the found spots from the average correct predictions to those reported in the studies, the spot from -666 ms to -466 ms is as expected. However, the other spot from 1733 ms to 2066 ms is outside the range where the studies expect important features to be seen. This suggests that the LSTM CNN is focusing on the temporal spot from 1733 ms to 2066 ms after movement onset, which according to Shakeel et al. [39], is when the muscles are going back to the relaxed state. Which could mean the LSTM CNN has found a feature right before the muscle relaxation to be important in the differentiation of the three classes.

When looking at Figure 3.6, which compares the attention weights from the average correct and incorrect predictions, a clear attention to the correct predictions can be seen from around -1000 ms to 1000 ms. This illustrates that the attention weights

are higher in value in the correct predictions compared to the incorrect predictions. Which is also what the studies Gervasio et al. [20], Sur and Sinha [38], Shakeel et al. [39], have found to be the important temporal time. But the figure also illustrates more attention to the correct predictions at the start and end of the signal. Which is, as mentioned in the study Shakeel et al. [39], seen as the muscle relaxation period. Therefore, it could mean that some information about the pain class, which is not yet researched in the literature, can be seen in the relaxation period both before and after the movement initiation.

Therefore, a future improvement of the LSTM CNN could be an implementation of a frequency analysis in the specific time intervals highlighted by the heatmap. This would provide a deeper insight into the features used for the classification, revealing which frequencies the model is focusing on in those highlighted temporal windows. Such improvement could enhance the explainability and potentially guide future pain studies by identifying the frequencies used for differentiation of the classes. And potentially explain why the model is focusing on the muscle relaxation spots, as potentially important for the classification.

4.5 Challenges in implementation and system limitations

The LSTM CNN was developed with the aim of helping differentiate the acute and chronic pain classes by the use of the EEG. Such technology would be categorised as a predictive technology. This means the system would be used by a general practitioner or at the primary line of care to differentiate the acute pain patients from the chronic pain patients. The LSTM CNN demonstrated this with an accuracy of 75.5 % and 84.9 % correct predictions, but due to the low healthy correct prediction rate, it would mean the current system should only be used in the situation of differentiating between the two pain states. Furthermore, the B, illustrated the relatively high Recall and Precision values for the two pain classes, except for the precision value in the acute pain class. This reflects that the system demonstrates low false negative and false positive cases, which is what is expected of a system for the healthcare industry.
For the chronic class, the Recall and Precision values were 0.76 and 0.81, and for the acute pain class, the Precision and Recall were 0.47 and 0.85. Which for both classes illustrates low False negative and low false positive predictions, except for the Precision of the acute class.
The precision value from the acute pain class was 0.47, which means there were multiple false positive predictions made. Thus, in this case, it was the healthy class that 75.2 % of the time was incorrectly predicted as acute pain, making the false positive value of the acute pain class rise.
Therefore, taking into account that the system should only differentiate between the two pain states, it would be suitable as a potential predictive technology.

Another limitation of the LSTM CNN was seen in the training and validation loss figure in Figure 3.1. This illustrates that there was a small overfitting happening during the training of the model. This can be seen as the training loss shows a steady decrease over time, compared to the validation loss, which flattens out slightly after 25 epochs, this can be seen in the left side of Figure 3.1. Furthermore, some spikes in the validation loss can, e.g. be seen in epochs 17 and 36, which could be the result of either a small validation set or noise in the validation data segments. Thus, when looking at the accuracy of the training and validation, seen on the right side of Figure 3.1, it does not show any major overfitting indications, but can also verify that epoch 17 might indicate some noise in the data segments of the validation set. This would therefore mean, for a better overall ML model, with the filters for the segments mentioned earlier in mind, a bigger dataset for the machine learning would minimise the model's chance of overfitting on the data, and potentially make the accuracy of the model better.

5 | Conclusion

In this project, the problem area of objective pain classification and differentiation between chronic, healthy, and acute pain was explored. This was done with the development of a ML model able to analyse EEG data segments and classify the segments into the three groups, while displaying where in the signal the focus was during the analysis and extraction of the features used for the classification.

The developed LSTM CNN achieved an average prediction accuracy of 54.55 %, with a class specific prediction accuracy of: Chronic pain accuracy of 75.68 %, healthy accuracy of 13.76 %, and acute pain accuracy of 84.95 %. As an output of the LSTM CNN, heatmaps of the attention weights displaying the area of focus, could be seen for the entire analysis, class specific, and solo segments. Illustrating where the features were extracted for the classes and some of the data segments.

Although the model demonstrated a relatively low overall accuracy, the F1 score indicated that it could learn meaningful patterns for the classification of the different classes. Additionally, the heatmap visualisations supported the findings from existing literature, with what important features were in the area of interest, and revealed a potential new classification feature related to the muscle relaxation time, which could be an interesting topic for future pain related movement studies.

One of the primary limitations of the current system was the small dataset, which led to a slight overfitting during the training of the LSTM CNN. Furthermore, the data contained noisy segments due to sparse preprocessing on the EEG. As a future improvement, training on more data and including a preprocessing part to the EEG, and therefore only including the important frequency bands, thereby excluding the noise. This could enhance the model's performance and potentially improve classification accuracy, especially for the healthy class.

Another promising direction for future improvement would be to further develop the LSTM CNN model by implementing a frequency analysis of the segments with the high attention weights from the heatmap. This could provide a more detailed insight into the relevant frequencies at those highlighted spots, potentially boosting classification performance and the explainability of the features used for the classification.

The LSTM CNN, therefore, proves its potential as a predictive technology by analysing objective EEG pain recordings and classifying them to differentiate the chronic pain, healthy, and acute pain classes. However, future improvements are required of the current setup, to get a better classification accuracy of the healthy class and to obtain a larger dataset for building a more robust ML model.

6 | Time management

This section is about how a time plan was first set up and later adjusted to better fit the project.

Initial time plan

At the beginning of the project, a time plan was created to provide a professional overview of the time I had available, assisting with meetings and deadlines, and keeping an overview of the tasks which needed to be completed throughout the writing process.
However, one problem with the initial time plan was that it was based on the workflow of group projects, which was what I had been used to up until this point, and also meant the amount of resources for the project was limited to now one person instead of three or five persons, therefore everything took longer time.

Therefore, I knew the initial time plan needed to be adjusted as the work progressed, but always with in mind that the time adjusted had to be taken from another task. The initial time plan was set up in the Microsoft Teams' Planner application, as seen in 6.1. Where the different tasks which were planned can be seen, and with a check mark indicating that the tasks are completed.

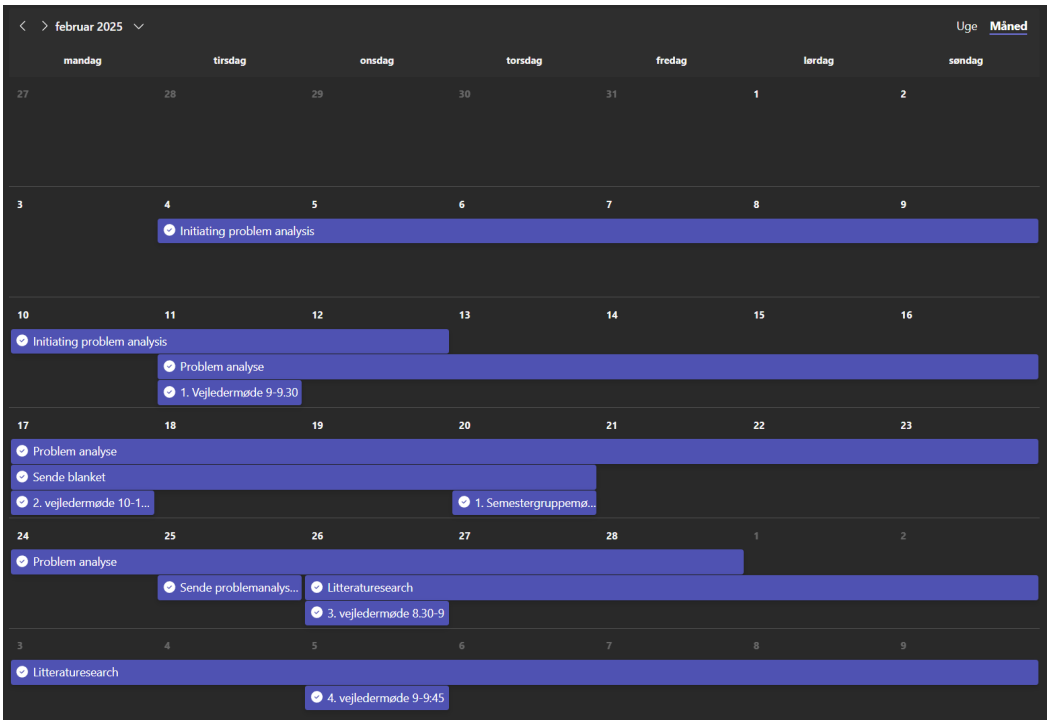

Figure 6.1: *Illustrates the initial time plan, which includes the different tasks and meetings during February.*

Adjustments to the time plan

Working solo on the project for the first two months led to a clearer understanding of the new workflow and time allocation for the different tasks.
Which resulted in a second look at the time plan, with the work done and the different upcoming deadlines in mind.
The new time plan was created using the backcasting method. With this approach, the final tasks and deadlines were first identified, such as the submission date, and then worked backwards to determine when each of the earlier tasks needed to be completed. Thereby, the different tasks were identified and set up with a deadline for completion or when a section needed to be sent for revision.
This was done with two different deadline types, which were hard deadlines and soft deadlines.[46] A hard deadline is a deadline which can not be postponed, such as this project's submission date. Where soft deadlines can be postponed, which could e.g. be finishing the code to proceed to the testing phase of the system.
By marking the different tasks with the deadline types, and including more reading and correction time when receiving revised text from the supervisor, it made the project more focused on solo project writing. Things that changed were, e.g. a longer time was reserved for writing and researching material, as it could not be split among different group members and then later shared the knowledge gained.

As a risk management part of ensuring the project was delivered on time, buffer periods were included in the time plan to account for potential sick days or personal issues. For example, an extra days were scheduled during the final phase of reading and adjusting the project. This buffer time was used, as some of the soft deadlines were postponed and adjusted by using the buffer time instead of taking the time from other tasks.
Furthermore, to ensure the time plan was kept and to maintain an overview of what tasks needed to be started or completed, the time plan was visited every Friday. This involved planning the upcoming week by updating the tasks in Microsoft Planner with notes on smaller subtasks which needed to be done. This weekly review helped keep focus on the current task and the upcoming deadlines.

With the new time plan, all hard deadlines were met, and only a few soft deadlines were postponed. This resulted in a more manageable and realistic schedule, which ultimately increased motivation, as the workload felt more achievable within the time available for the project. Therefore, in the end, the new time plan and the method used to make it showed the importance of knowing how to manage the time available and the importance of revisiting the time plan during the process of making a project like this.

Bibliography

[1] World Health Organization: Who. Musculoskeletal health. *World Health Organization: WHO*, July 2022. URL `https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions`.

[2] Daniel S. Goldberg and Summer J. McGee. Pain as a global public health priority. *BMC Public Health*, 11:770, October 2011. doi:10.1186/1471-2458-11-770.

[3] Salah N. El-Tallawy, Rohit Nalamasu, Gehan I. Salem, Jo Ann K. LeQuang, et al. Management of Musculoskeletal Pain: An Update with Emphasis on Chronic Musculoskeletal Pain. *Pain Ther.*, 10(1):181–209, June 2021. ISSN 2193-651X. doi:10.1007/s40122-021-00235-2.

[4] Jinlei Zhou, Shanggao Xie, Sen Xu, Yuan Zhang, Yanlei Li, Qice Sun, Jun Zhang, and Tingxiao Zhao. From pain to progress: Comprehensive analysis of musculoskeletal disorders worldwide. *Journal of Pain Research*, pages 3455–3472, 2024.

[5] International Association for the Study of Pain (IASP). Iasp revised definition of pain, 2022. URL `https://www.iasp-pain.org/wp-content/uploads/2022/04/revised-definition-flysheet_R2-1-1-1.pdf`. [Online; accessed 19. Feb. 2025].

[6] Joanna Bielewicz, Maciej Kamieniak, Michał Szymoniuk, Jakub Litak, et al. Diagnosis and Management of Neuropathic Pain in Spine Diseases. *Journal of Clinical Medicine*, 12(4):1380, February 2023. doi:10.3390/jcm12041380.

[7] Xiang-Yao Li, You Wan, Shao-Jun Tang, Yun Guan, et al. Maladaptive Plasticity and Neuropathic Pain. *Neural Plast.*, 2016:4842159, January 2016. doi:10.1155/2016/4842159.

[8] Rohini Kuner and Herta Flor. Structural plasticity and reorganisation in chronic pain. *Nature Reviews Neuroscience*, 18(1):20–30, 2017.

[9] Eva Kosek, Daniel Clauw, Jo Nijs, Ralf Baron, et al. Chronic nociplastic pain affecting the musculoskeletal system: clinical criteria and grading system. *Pain*, 162(11):2629, June 2021. ISSN 0304-3959. doi:10.1097/j.pain.0000000000002324.

[10] Jing Luo, Hui-Qi Zhu, Bo Gou, and Xue-Qiang Wang. Neuroimaging Assessment of Pain. *Neurotherapeutics*, 19(5):1467–1488, September 2022. ISSN 1878-7479. doi:10.1007/s13311-022-01274-z.

[11] OpenStax. Anatomy and physiology – "central processing". https://openstax.org/books/anatomy-and-physiology/pages/1-introduction, 2025. Access for free at https://openstax.org/books/anatomy-and-physiology/pages/1-introduction.

[12] Leonard Verhagen Katie Kompoliti. Event-related potentials: Slow potentials. In Semyon Slobounov, editor, *Encyclopedia of Movement Disorders*, pages 456–458. Elsevier's online platform of Science Direct, 2010. doi:10.1016/B978-0-12-374105-9.00458-5. [Online; accessed 19. Feb. 2025].

[13] Jesper Bie Larsen, Pernille Borregaard, Janus Laust Thomsen, Michael Skovdal Rathleff, and Simon Kristoffer Johansen. Improving general practice management of patients with chronic musculoskeletal pain: Interdisciplinarity, coherence, and concerns. *Scandinavian Journal of Pain*, 24(1):20230070, 2024.

[14] American Society of Anesthesiologists Task Force on Chronic Pain Management et al. Practice guidelines for chronic pain management: an updated report by the american society of anesthesiologists task force on chronic pain management and the american society of regional anesthesia and pain medicine. *Anesthesiology*, 112(4):810–833, 2010.

[15] Jeremy Lewis and Peter O'Sullivan. Is it time to reframe how we care for people with non-traumatic musculoskeletal pain?, 2018.

[16] Kimberly T Sibille, Felix Bartsch, Divya Reddy, Roger B Fillingim, and Andreas Keil. Increasing neuroplasticity to bolster chronic pain treatment: a role for intermittent fasting and glucose administration? *The journal of pain*, 17(3):275–281, 2016.

[17] Panagiotis Zis, Andreas Liampas, Artemios Artemiadis, Gabriela Tsalamandris, et al. EEG Recordings as Biomarkers of Pain Perception: Where Do We Stand and Where to Go? *Pain Ther.*, 11(2):369–380, June 2022. ISSN 2193-651X. doi:10.1007/s40122-022-00372-2.

[18] Zhenyuan Lu, Burcu Ozek, and Sagar Kamarthi. Transformer encoder with multiscale deep learning for pain classification using physiological signals. *Front. Physiol.*, 14:1294577, December 2023. ISSN 1664-042X. doi:10.3389/fphys.2023.1294577.

[19] Mika M Rockholt, George Kenefati, Lisa V Doan, Zhe Sage Chen, and Jing Wang. In search of a composite biomarker for chronic pain by way of eeg and machine learning: where do we currently stand? *Frontiers in neuroscience*, 17: 1186418, 2023.

[20] Sabata Gervasio, Ali Asghar Zarei, and Natalie Mrachacz-Kersting. Eeg signatures of low back and knee joint pain during movement execution: a short report. *Frontiers in Rehabilitation Sciences*, 4:1216069, 2023.

[21] Ian Gilron, Troels S Jensen, and Anthony H Dickenson. Combination pharmacotherapy for management of chronic pain: from bench to bedside. *The Lancet Neurology*, 12(11):1084–1095, 2013.

[22] Dong Hee Lee, Sungwoo Lee, and Choong-Wan Woo. Decoding pain: uncovering the factors that affect the performance of neuroimaging-based pain models. *Pain*, 166(2):360–375, 2025.

[23] Xiaohan Xu and Yuguang Huang. Objective pain assessment: a key for the management of chronic pain. *F1000Research*, 9:F1000–Faculty, 2020.

[24] Li-Bo Zhang, Yu-Xin Chen, Zhen-Jiang Li, Xin-Yi Geng, Xiang-Yue Zhao, Feng-Rui Zhang, Yan-Zhi Bi, Xue-Jing Lu, and Li Hu. Advances and challenges in neuroimaging-based pain biomarkers. *Cell Reports Medicine*, 2024.

[25] Vishal Vijayakumar, Michelle Case, Sina Shirinpour, and Bin He. Quantifying and characterizing tonic thermal pain across subjects from eeg data using random forest models. *IEEE Transactions on Biomedical Engineering*, 64(12): 2988–2996, 2017.

[26] Katherine T Martucci and Sean C Mackey. Neuroimaging of pain: human evidence and clinical relevance of central nervous system processes and modulation. *Anesthesiology*, 128(6):1241, 2018.

[27] Younbyoung Chae, Hi-Joon Park, and In-Seon Lee. Pain modalities in the body and brain: Current knowledge and future perspectives. *Neuroscience & Biobehavioral Reviews*, 139:104744, 2022.

[28] Maite M Van Der Miesen, Martin A Lindquist, and Tor D Wager. Neuroimaging-based biomarkers for pain: state of the field and current directions. *Pain reports*, 4(4):e751, 2019.

[29] Colince Meli Segning, Hassan Ezzaidi, Rubens A da Silva, and Suzy Ngomo. A neurophysiological pattern as a precursor of work-related musculoskeletal disorders using eeg combined with emg. *International Journal of Environmental Research and Public Health*, 18(4):2001, 2021.

[30] Jerin Mathew, Tyson Michael Perez, Divya Bharatkumar Adhia, Dirk De Ridder, and Ramakrishnan Mani. Is there a difference in eeg characteristics in acute, chronic, and experimentally induced musculoskeletal pain states? a systematic review. *Clinical EEG and neuroscience*, 55(1):101–120, 2024.

[31] Jonathan Miller, Skylar Jacobs, William Koppes, Frank Minella, Federica Porta, Fletcher A White, and Joseph A Lovelace. Development of machine learning algorithms using eeg data to detect the presence of chronic pain. *medRxiv*, pages 2024–09, 2024.

[32] Hyeon Seok Seok, Sang Su Kim, Do-Won Kim, and Hangsik Shin. Toward objectification of subjective chronic pain based on implicit response in biosignals. *IEEE Transactions on Biomedical Engineering*, 2024.

[33] Colince Meli Segning, Rubens A da Silva, and Suzy Ngomo. An innovative eeg-based pain identification and quantification: A pilot study. *Sensors*, 24(12): 3873, 2024.

[34] J Tsigarides, A Rushbrooke, and A Bagnall. Pos0257 painwaves: The potential of machine learning to differentiate chronic pain cohorts using electroencephalography. *Annals of the Rheumatic Diseases*, 82:366, 2023.

[35] Jeffrey W. Britton, Lauren C. Frey, Jennifer L. Hopp, Pearce Korb, Mohamad Z. Koubeissi, William E. Lievens, Elia M. Pestana-Knight, and Erik K. St. Louis. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. American Epilepsy Society, 2016. ISBN 9780997975604.

[36] Timo Kirschstein and Rüdiger Köhling. What is the source of the eeg? *Clinical EEG and neuroscience*, 40(3):146–149, 2009.

[37] Brandon J Thio and Warren M Grill. Relative contributions of different neural sources to the eeg. *Neuroimage*, 275:120179, 2023.

[38] Shravani Sur and Vinod Kumar Sinha. Event-related potential: An overview. *Industrial psychiatry journal*, 18(1):70–73, 2009.

[39] Aqsa Shakeel, Muhammad Samran Navid, Muhammad Nabeel Anwar, Suleman Mazhar, Mads Jochumsen, and Imran Khan Niazi. A review of techniques for detection of movement intention using movement-related cortical potentials. *Computational and mathematical methods in medicine*, 2015(1):346217, 2015.

[40] Salvador Daniel Rivas-Carrillo, Evgeny E Akkuratov, Hector Valdez Ruvalcaba, Angel Vargas-Sanchez, Jan Komorowski, Daniel San-Juan, and Manfred G Grabherr. Mindreader: Unsupervised classification of electroencephalographic data. *Sensors*, 23(6):2971, 2023.

[41] Tong Zhao, Yi Cui, Taoyun Ji, Jiejian Luo, Wenling Li, Jun Jiang, Zaifen Gao, Wenguang Hu, Yuxiang Yan, Yuwu Jiang, et al. Vaeeg: Variational auto-encoder for extracting eeg representation. *NeuroImage*, 304:120946, 2024.

[42] Swati Rajwal and Swati Aggarwal. Convolutional neural network-based eeg signal analysis: A systematic review. *Archives of Computational Methods in Engineering*, 30(6):3585–3615, 2023.

[43] Lipnitskii Mikhail. Meta-analysis of eeg findings on pain perception: Exploring nociceptive and neuropathic pain response patterns. *bioRxiv*, pages 2023–10, 2023.

[44] Paul W Hodges and Bang H Bui. A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography. *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, 101(6):511–519, 1996.

[45] Yuxin Yan, Haifeng Zhou, Lixin Huang, Xiao Cheng, and Shaolong Kuang. A novel two-stage refine filtering method for eeg-based motor imagery classification. *Frontiers in neuroscience*, 15:657540, 2021.

[46] Effectiviology. Deadlines: How effective time constraints can boost productivity. `https://effectiviology.com/deadlines/`, 2025. Access for free at `https://effectiviology.com/`.

Appendix

A | Heatmaps of the solo segments

This abstract includes some of the results from the LSTM CNN, which is the solo segments and the prediction made by the ML model.
The following segments are displayed with the EEG segments at the top of the figure, and the heatmap of what the LSTM CNN focused on when classifying the data segment. Furthermore, the prediction and the correct class can be seen as the title of the figure, with the colour of the tile indicating if the prediction was correct (green title colour) or incorrect (red title colour).

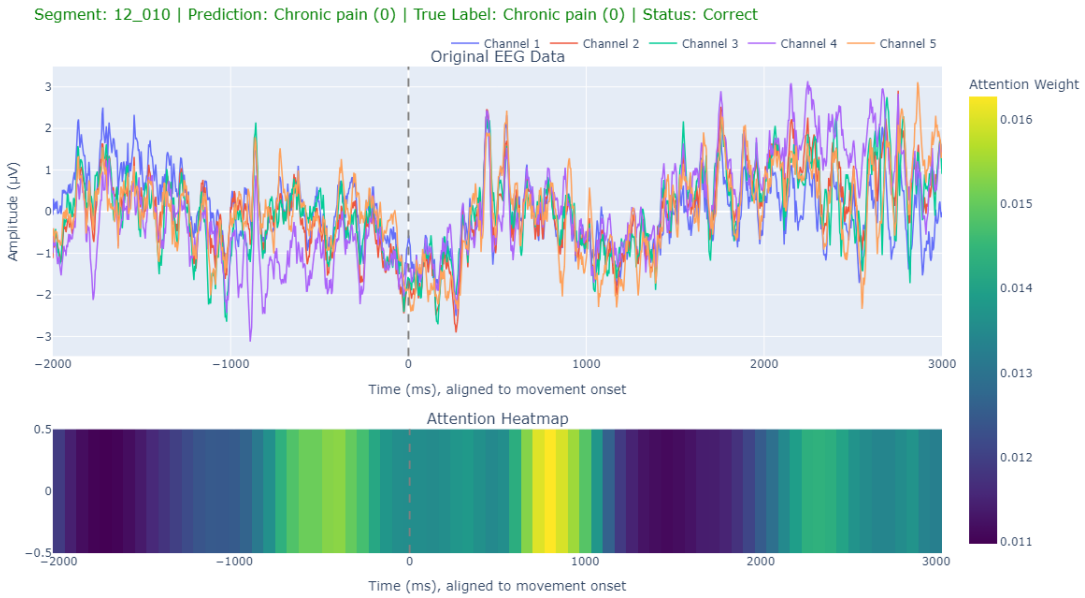

Figure A.1: *Illustrates an EEG segment from a chronic pain participant, which was correctly predicted. The top part displays EEG signals from the 5 channels, while the bottom part shows the corresponding attention heatmap. Both are aligned to the movement onset at 0 ms on the x-axis.*

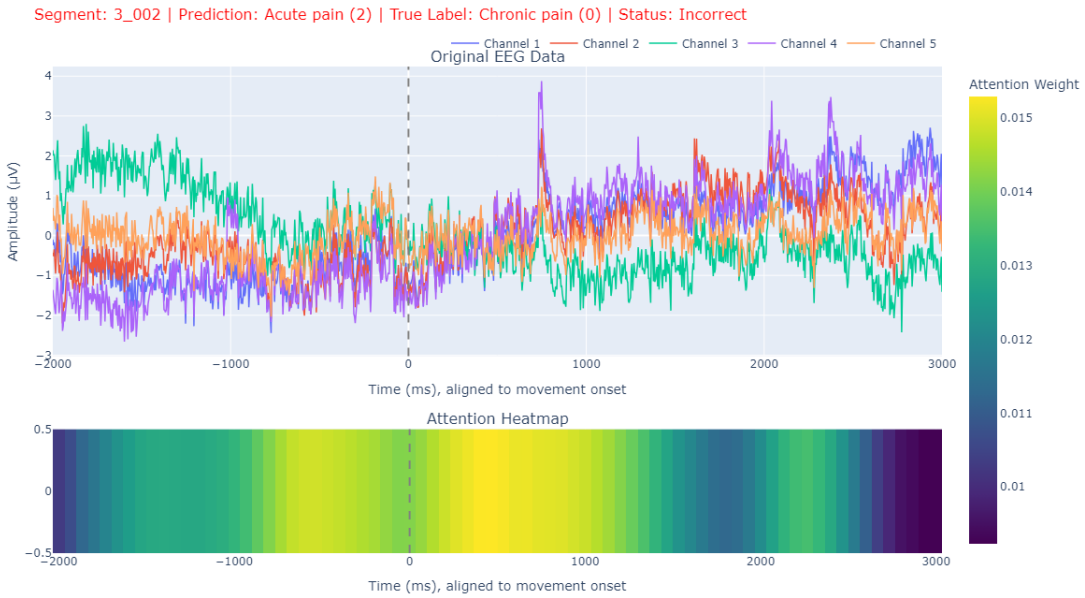

Figure A.2: *Illustrates an EEG segment from a participant with chronic pain, which was incorrectly classified as acute pain. The top part displays EEG signals from the 5 channels, while the bottom part shows the corresponding attention heatmap. Both are aligned to the movement onset at 0 ms on the x-axis.*

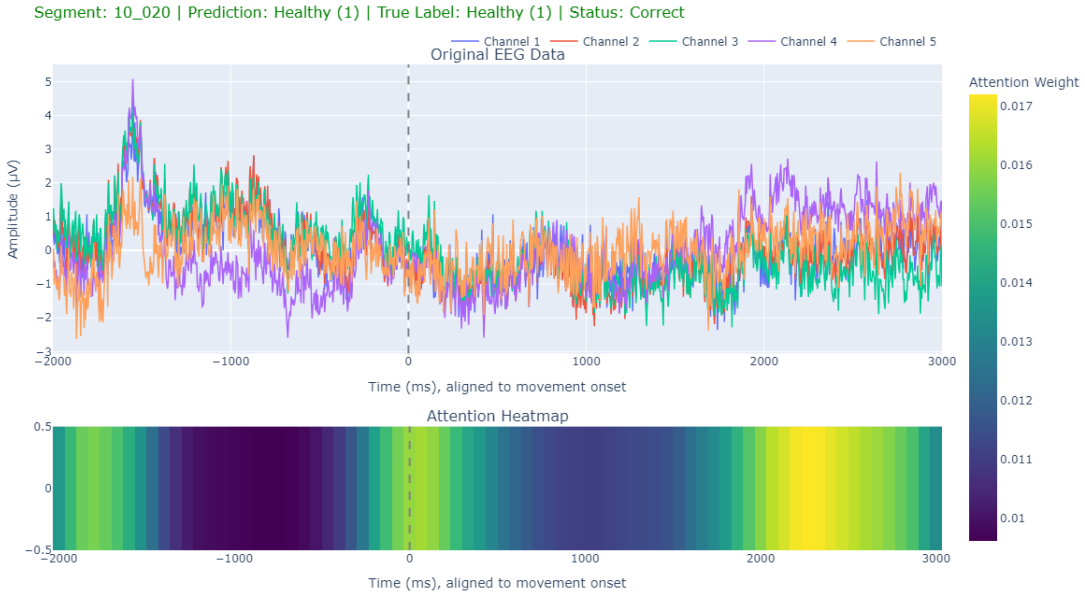

Figure A.3: *Illustrates an EEG segment from a healthy participant, which was correctly predicted. The top part displays EEG signals from the 5 channels, while the bottom part shows the corresponding attention heatmap. Both are aligned to the movement onset at 0 ms on the x-axis.*

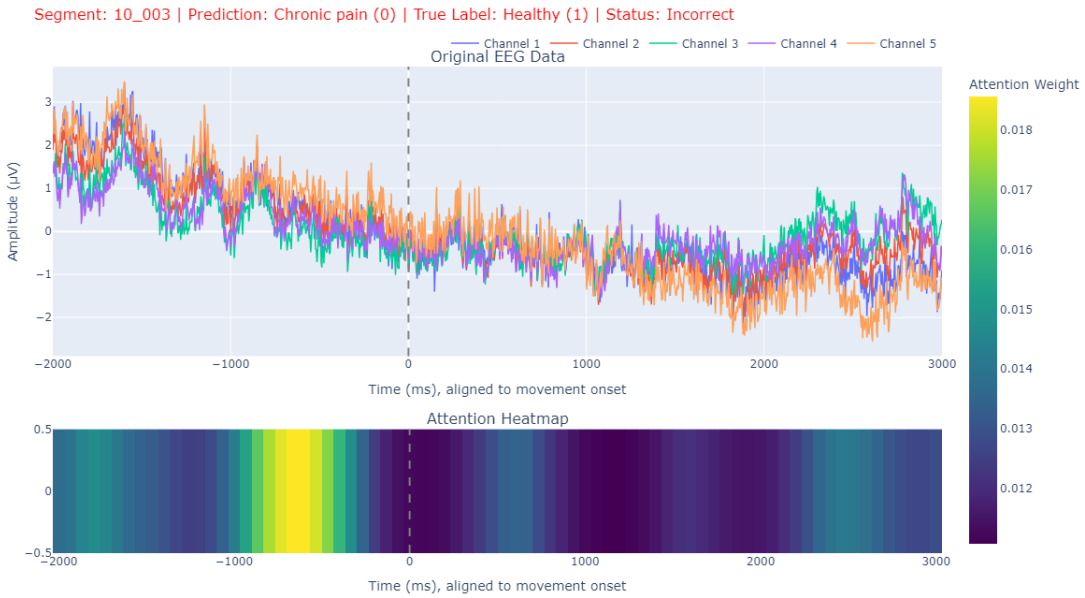

Figure A.4: *Illustrates an EEG segment from a healthy participant, which was incorrectly classified as chronic pain. The top part displays EEG signals from the 5 channels, while the bottom part shows the corresponding attention heatmap. Both are aligned to the movement onset at 0 ms on the x-axis.*

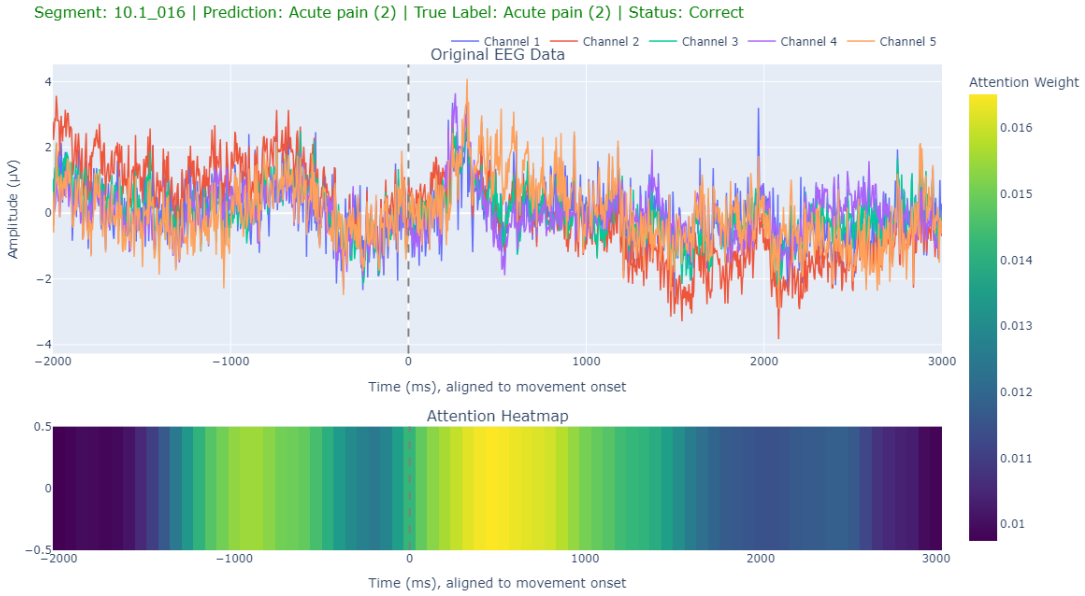

Figure A.5: *Illustrates an EEG segment from an acute participant, which was correctly predicted. The top part displays EEG signals from the 5 channels, while the bottom part shows the corresponding attention heatmap. Both are aligned to the movement onset at 0 ms on the x-axis.*

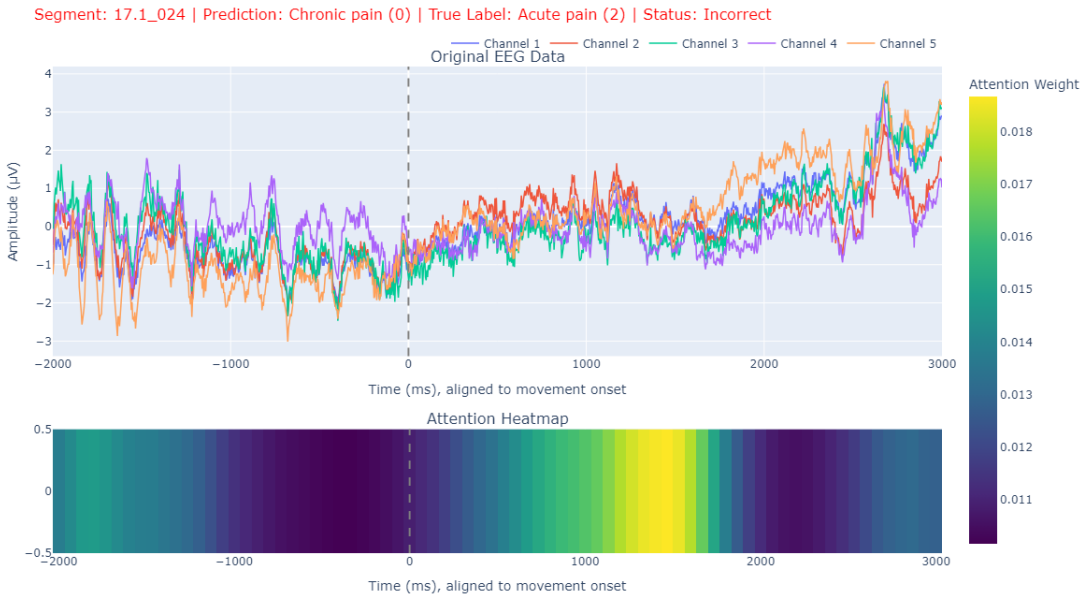

Figure A.6: *Illustrates an EEG segment from an acute pain participant, which was incorrectly classified as chronic pain. The top part displays EEG signals from the 5 channels, while the bottom part shows the corresponding attention heatmap. Both are aligned to the movement onset at 0 ms on the x-axis.*

B | Calculation and formulas for the confusion matrix.

This abstract includes the calculations made on the confusion matrix, including the formulas which was used.
The different calculations done on the confusion matrix can be seen with the results. The first thing to identify is the value of the True positive, True negative, False positive, and False negative in the confusion matrix. This has been done, and the values can be seen in Table B.1.

Table B.1: *Illustrates the values extracted from the confusion matrix. The classes of which the values are calculated can be seen on the left. In the table, some abbreviations can be seen at the top, which are: TP = True positive, TN = True negative, FP = False positive, FN = False negative.*

	TP	TN	FP	FN
Chronic	56	189	13	18
Healthy	15	143	24	94
Acute	79	94	89	14

After the class specific values have been calculated, the values can be used to calculate the class specific Precision and class specific Recall values. The class specific Precision formula is:

$$Precision : \frac{TP}{FP + TP}$$

The class specific recall formula is:

$$Recall : \frac{TP}{FN + TP}$$

The precision value describes the system's capability to predict True positive cases, as more false positive cases would decrease the value of the final result of the calculated Precision. A perfect Precision value would therefore be 1, as this means the system only predicts True positive cases. The Recall value reflects the system's ability to not predict the data as False negative.

The results from each class can be seen in Table B.2, where the formulas are used to calculate the class specific values.

Table B.2: *Shows the calculated class specific Precision and Recall values.*

	Precision	Recall
Chronic	0.812	0.757
Healthy	0.385	0.138
Acute	0.47	0.849

These results of the class specific Recall and Precision values give information about how well the LSTM CNN was at classifying each class.
The values from the Table B.1 are used to make the Micro Recall and Micro Precision calculations, which are the average Recall and Precision values of the confusion matrix, when the dataset is unbalanced, as this test set is. The Micro Recall and Micro Precision values are calculated with the True positive and False positive from each class.

$$Precision_{Micro} : \frac{TP_{Chronic} + TP_{Healthy} + TP_{Acute}}{TP_{Chronic} + FP_{Chronic} + TP_{Healthy} + FP_{Healthy} + TP_{Acute} + FP_{Acute}}$$

$$Recall_{Micro} : \frac{TP_{Chronic} + TP_{Healthy} + TP_{Acute}}{TP_{Chronic} + FN_{Chronic} + TP_{Healthy} + FN_{Healthy} + TP_{Acute} + FN_{Acute}}$$

The result of the Micro Precision was calculated to 0.543, and the micro Recall value was calculated to 0.543. The Micro Precision and Micro Recall are both used for the calculation of the F1 score, which can be seen in the following formula.

$$F1Score : \frac{2 * Precision_{Micro} * Recall_{Micro}}{Precision_{Micro} + Recall_{Micro}}$$

The result of the F1 score was calculated to 0.543, which is classified as an ok model.