**1.1 Исследование биофизических основ формирования движения**

Желание напрячь какую-либо мышцу зарождается в головном мозге человека. Этот сигнал, представляющий собой электрохимический импульс, переносится между нейронами – узкоспециализированными клетками, являющимися структурно-функциональными единицами нервной системы человека. Нейроны, присоединенные непосредственно к мышечным клеткам, называются мотонейронами. Совокупность мотонейрона и всех мышечных волокон, иннервируемых им, называется двигательной единицей (ДЕ).

Крупные мышцы, служащие для выполнения больших движений, насчитывают сотни волокон в каждой ДЕ; мышцы для точных движений имеют меньшее количество волокон в каждой единице. Количество мышечных волокон, приходящихся на одно двигательное нервное волокно, известно как коэффициент иннервации. При возбуждении двигательной единицы происходит распространение потенциала действия – разности электрических потенциалов относительно невозбуждённой клетки. Механизм распространения потенциала действия имеет важный характер при рассмотрении электрической активности мышечной ткани.

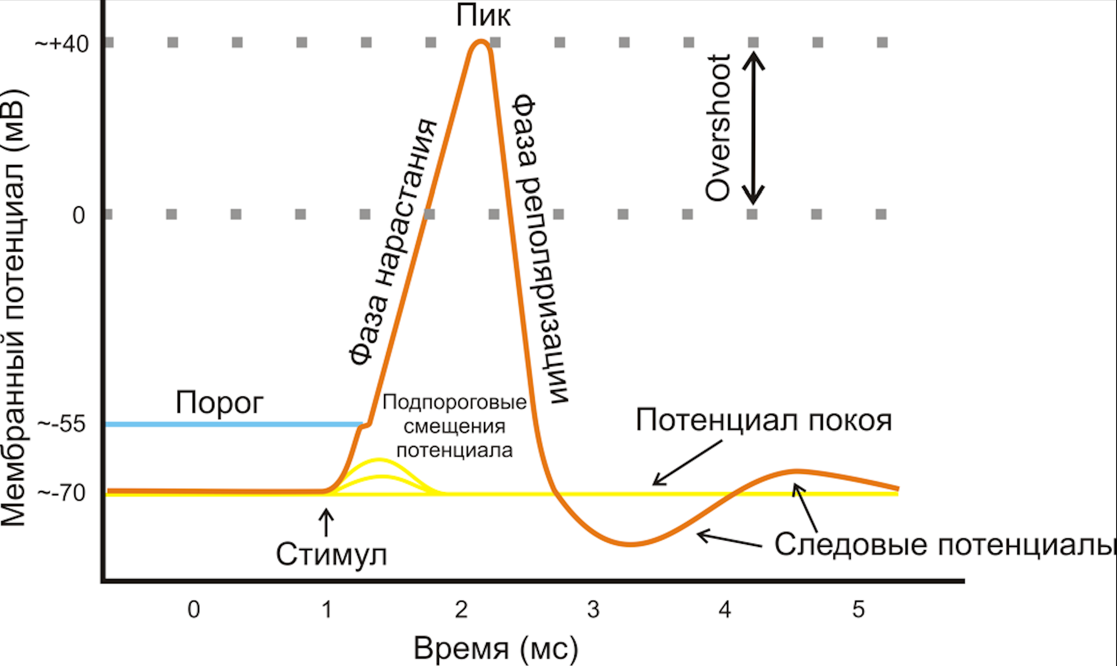
Обычно в потенциале действия выделяют четыре фазы: фаза покоя, фаза деполяризации, фаза реполяризации и фаза гиперполяризации. Изначально нейрон находится в покое и имеет стабильный отрицательный внутриклеточный потенциал. Затем происходит деполяризация, обусловленная работой «натрий-калиевого насоса». Концентрация ионов натрия в цитоплазме клетки в невозбуждённом состоянии очень мала по сравнению с окружающей средой. При проведении потенциала действия открываются потенциал-зависимые натриевые каналы и положительно заряженные ионы натрия поступают в цитоплазму по градиенту концентрации, пока он не будет уравновешен положительным электрическим зарядом. Вслед за этим происходит реполяризация: потенциал-зависимые каналы инактивируются и отрицательный потенциал покоя восстанавливается за счёт диффузии из клетки положительно заряженных ионов калия, концентрация которых в окружающей среде значительно ниже внутриклеточной. Наконец, наступает фаза гиперполяризации. Мембрана нейрона становится временно более проницаемой для калиевых ионов, что приводит к снижению электрического потенциала внутри клетки ниже базового уровня покоя. Это состояние создает рефрактерный период, когда нейрон не может быть возбужден новым потенциалом действия. В норме ПДОДЕ обычно имеют длительность 3-15 мс, амплитуду 100-300 мкВ и появляются с частотой в пределах от 6 до 30 раз в секунду [1]. Типичный вид потенциала действия изображен на рисунке 1.

Рис. 1. Схематичное изображение идеализированного потенциала действия.

При стимуляции нейронным сигналом каждая двигательная единица сокращается и вызывает электрический сигнал, который представляет собой сумму потенциалов действия всех вовлечённых в процесс клеток. Этот потенциал известен как потенциал действия отдельной двигательной единицы (ПДОДЕ или просто ПДДЕ). На рисунке 2 схематично изображена двигательная единица, а также этапы формирования миоэлектрического сигнала. Закрашенные мышечные волокна иллюстрируют иннервируемые данным мотонейроном волокна. Импульсы возбуждения представлены в виде дельта-функций, – функции, иллюстрирующие генерацию ПДОДЕ i-той ДЕ. Снизу изображено влияние шумов регистрирующей аппаратуры на миоэлектрический сигнал.

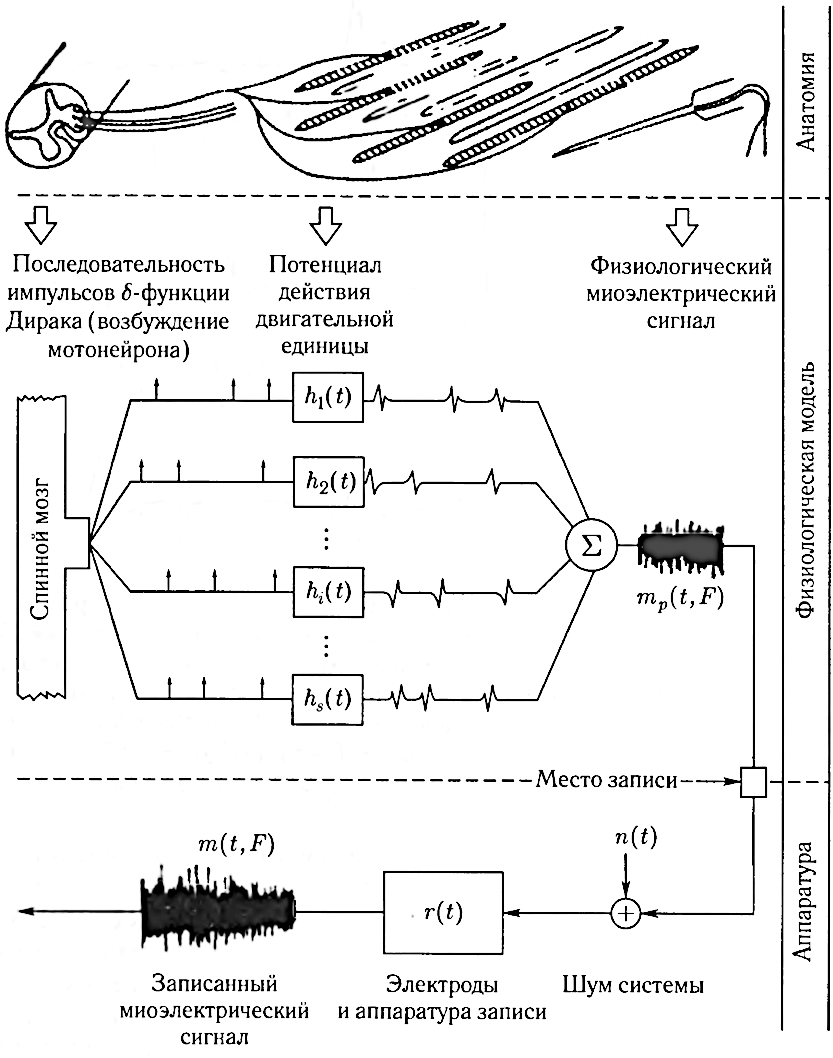


Рисунок 2: Схематическое представление ДЕ и модели генерации миоэлектрического сигнала.

**1.2 Изучение способов оценки электрической активности мышц**

Дословно термин электромиография можно расшифровать как запись электрической активности мышц. Существует два основных метода регистрации ЭМГ сигнала: инвазивный и неинвазивный (поверхностный). При инвазивном методе используют игольчатые электроды, вживляемые в мышечную ткань. При поверхностном методе используют электроды, располагаемые на коже. Достоинством инвазивного метода является большая точность и большее отношение сигнал/шум, поскольку сигнал считывается непосредственно с мышечного волокна. Главным недостатком является неудобство использования (необходимость вживления игольчатых электродов под кожу человека). Анализ многочисленных исследований в области распознавания жестов и движений по поверхностной ЭМГ (пЭМГ) подтверждает целесообразность использования именно этого метода.

Для регистрации сигнала пЭМГ используется несколько видов электродов. Существует две основные группы: сухие и мокрые пЭМГ электроды. Мокрые электроды, как правило, обладают лучшими регистрирующими свойствами, но имеют целый ряд недостатков: при их использовании необходима подготовка места контакта с электродом – сбривание волос, «зачистка кожи абразивом». Также, как следует из названия, для их работы необходим специальный гель. Гель может приводить к аллергической реакции, раздражению кожи, росту бактерий. Помимо прочего, при использовании мокрых электродов, необходимо использовать клейкое крепление. Оно также способно приводить к раздражению и аллергии. Кожа под клеем может вспотеть, что может привести к ухудшению качества крепления и деградации получаемого сигнала. Наконец, при длительном использовании мокрого электрода, происходит его обезвоживание и качество регистрируемого сигнала ухудшается [2]. Исходя из вышеперечисленных недостатков, более простым и практичным будет использование сухих электродов. Сухие электроды различаются по материалам, из которых изготовлены, форме и площади. Электроды, изготовленные из оксида графена или углеродных нанотрубок, обладают наилучшими характеристиками по качеству сигнала и удобству использования. Однако, из-за высокой стоимости углеродных наноматериалов, на сегодняшний день данная технология распространена слабо. Наиболее доступными являются электроды, изготовленные из медицинской стали или других металлических сплавов. Крепить такие электроды можно с помощью компрессионной ткани либо используя плотно прилегающий браслет.

**1.3 Установление особенностей, влияющих на изменчивость сигнала**

Существует целый ряд особенностей, влияющих на пЭМГ сигнал. Авторы статьи [3] разделили их на две большие группы – физиологические и нефизиологические. К физиологическим факторам относятся: свойства волоконной мембраны (средняя скорость проводимости мышечных волокон, распределение скорости проводимости мышечных волокон, форма внутриклеточных ПД и др.) и свойства ДЕ (количество задействованных ДЕ, статистика и коэффициент вариации для скорости разряда ДЕ и др.). Физиологические факторы определяются индивидуальными особенностями мышечной физиологии человека и не поддаются вариации при снятии пЭМГ.

Вторая группа факторов – нефизиологические, к ним относят: анатомические (форма объёмного проводника, коим является конкретная мышца; толщина подкожной клетчатки; тканевые неоднородности; длина мышечных волокон и другие анатомические особенности), геометрические (укорочение мышечных волокон при не изометрическом сокращении; смещение мышцы относительно электрода при её сокращении), физические (электропроводность тканей и количество перекрестных помех от ближайших мышц) и, наконец, факторы, определяемые регистрирующей системой (контакт кожа-электрод, фильтрация регистрируемого сигнала, расстояние между электродами, форма и площадь электродов, расположение электродов). Понятно, что для улучшения качества пЭМГ сигнала проще всего варьировать параметры регистрирующей системы. Оказывается, что при правильно подобранном положении электродов удаётся нивелировать влияние многих нефизиологических факторов. По этой причине при снятии пЭМГ всегда следует указывать положение электродов.

Одной из особенностей сигнала пЭМГ является сильное влияние перекрестных помех. Перекрестные помехи относят к сигналу, записанному на одной мышце, источником которого на самом деле является соседняя мышца, а передача электроды происходит через промежуточный объём. Учёт перекрестных помех является важным и сложным этапом обработки пЭМГ сигнала.

Помимо перечисленных факторов, искажающих сигнал, источником шума может служить внешнее электромагнитное излучение. Например, шум от электросети на частоте 50–60 Гц попадает в важный частотный диапазон максимума амплитуды пЭМГ сигнала 50–100 Гц [4]. Из-за большого количества искажающих факторов извлечение полезной информации из пЭМГ является затруднительным в основном из-за низкого отношения сигнал/шум (особенно при низкой мышечной активности, поскольку она бывает едва отличима от фоновой).

**1.4 Исследование подходов и признаков для классификации ЭМГ сигналов**

Сигнал пЭМГ является временным рядом. Из физических соображений понятно, что у этого ряда постоянный тренд (другими словами, математическое ожидание пЭМГ сигнала постоянно), конечная и постоянная дисперсия (амплитуда пЭМГ обычно не превышает несколько мВ). Временные ряды с постоянными статистическими характеристиками называют стационарными временными рядами. Распознавание жестов по пЭМГ сигналу по сути является задачей классификации. Существует множество способов создания классификатора для пЭМГ сигнала.

**** Классические методы заключаются в подборе и использовании наиболее значимых признаков из сигнала, содержащих достаточную для классификации информацию. Например, можно анализировать признаки в частотной области: коэффициенты авторегрессии, коэффициенты Фурье и вейвлет-коэффициенты. Более легко вычислить признаки во временной области. В статье [5] исследовались наиболее подходящие признаки во временной области, они перечислены в таблице 1.

Таблица 1: Временные функции для выделения признаков.

После выделения признаков авторы использовали для классификации нейронную сеть на основе радиальных базисных функций, главное преимущество которой заключается в способности обучаться в течение всего нескольких эпох. В итоге признак MAX показал наибольшую точность и скорость обучения.

В статье [6] для распознавания жестов рук по пЭМГ авторы использовали два линейных классификатора – байесовский и KNN (k-nearest neighbors; k-ближайших соседей). Для распознавания релевантных паттернов (начало и конец движения) подбирались индивидуальные пороговые значения (поскольку пЭМГ сигнал у каждого человека отличается). Другими словами, предварительно обработанное значение помечалось как начало шаблона, если было достигнуто определенное пороговое значение. Было обнаружено, что данный подход достаточно точен, поскольку существует значительная разница в амплитуде сигнала между пребыванием в исходном положении и началом выполнения жеста. Более сложным оказалось определение окончания паттерна. Если конечная граница будет слишком высокой, то паттерн будет пойман не полностью. Наоборот, если граница будет слишком близкой, то конец паттерна будет обнаружен поздно. Авторы пришли к выводу, что наиболее точным способом определения момента окончания паттерна является наблюдение за среднеквадратичным значением последних 16 входящих значений. Если это значение дважды подряд опускалось ниже двух граничных значений (возможно, различных), то это считалось концом шаблона. Затем происходило извлечение признаков, таких как уже знакомые максимум, минимум, среднее значение, дисперсия, длина сигнала и среднее квадратическое, а также признаки из частотной области (получаемые с помощью быстрого преобразования Фурье): фундаментальная частота, дисперсия Фурье спектра, длина части спектра, содержащей большие значения, чем среднее арифметическое коэффициентов Фурье (если в спектре несколько таких участков, их длины суммируются). Помимо вышеперечисленного, авторы включили особые функции, по-видимому, подходящие для различения четырех жестов: относительные позиции максимума и минимума в шаблоне и количество пересечений/прикосновений к нулевой линии отнесенное к длине сигнала. Как было сказано ранее, для создания классификатора, работающего в режиме реального времени, авторы использовали комбинацию двух быстрых алгоритмов: байесовский и KNN. Суть алгоритма KNN, как можно понять из названия, заключается в нахождении k-ближайших векторов выборок в n-мерном пространстве признаков. Близость определяется функцией расстояния, наиболее часто применяется евклидова метрика. К классу, содержащему наибольшее количество ближайших векторов из k-ближайших к данному, и будет отнесена данная выборка. Байесовский классификатор ­– это широкий класс алгоритмов, основанных на максимизации апостериорной (условной) вероятности. В простейшем случае байесовский классификатор может быть устроен следующим образом. Предполагается, что признаки, принадлежащие одному классу, подчиняются нормальному распределению и независимы. Тогда, используя теорему Байеса, можно показать, что вероятность того, что данная выборка X принадлежит некоему классу K, максимальна, когда максимально произведение вероятности события K на вероятность наблюдения X при условии того, что X принадлежит K. Класс, для которого это произведение окажется максимальным, и будет считаться содержащим выборку.

Применение в рассматриваемой задаче алгоритмов классификации, не использующих глубокое обучение, обладает неоспоримым преимуществом: как правило, они требуют от человека генерации меньшего количества данных.[7] Несмотря на это, в последние годы наблюдается экспоненциальный рост количества статей, в которых рассматриваются именно нейронные сети (причем в большинстве случаев (60%) рассматривается сверточная нейронная сеть).[8] Причина этого заключается в преимуществах нейронных сетей, которые будут перечислены далее.

Во-первых, нейронные сети способны самостоятельно извлекать наиболее подходящие признаки. Как было показано ранее, признаки, выделяемые вручную, являются достаточно простыми и потому не всегда оказываются наиболее достоверными при решении задачи классификации. Во-вторых, нейронные сети обладают большей универсальностью. Другими словами, нейросетевые алгоритмы могут быть применены к различным типам жестов и могут быть обучены для распознавания новых жестов без необходимости изменения алгоритма, в отличие от классических методов, которые могут требовать дополнительной настройки для каждого нового жеста. В третьих, благодаря автоматическому извлечению признаков, нейросетевые алгоритмы могут быть более устойчивы к шуму и изменениям, причины возникновения которых описаны в предыдущей главе. Наконец, нейронные сети показывают высокую точность при классификации пЭМГ сигналов. В ранее упомянутой статье [8] перечислены работы, в которых удалось добиться точности 80–90% при классификации десяти и более жестов.

**1.5 Исследование применения нейросетевых алгоритмов в задаче классификации движений человека.**

Интерес использования нейросетевых алгоритмов в задаче классификации движений человека по электромиографическим сигналам в первую очередь вызван высокой результативнотью применения этих алгоритмов во многих других сферах обработки биологической информации, таких как: диагностика болезней, прогнозирование заболеваний, обработка медицинских изображений, анализ медицинских текстов, анализ генетических данных, мониторинг пациентов и другие. В статье [9] авторы использовали CNN для создания классификатора жестов руки в режиме реального времени для управления роботизированным манипулятором. В качестве регистрирующего устройства использовался недорогой браслет Myo Armband (Thalmic Labs). Myo Armband имеет 8 сухих пЭМГ датчика с частотой дискретизации 200 Гц, а также трехосевые гироскопы и акселерометры. Для управления манипулятором были определены 7 интуитивно понятных жестов кисти руки: нейтральный, раскрытие кисти, сжатие кисти, сгибание запястья, разгибание запястья, лучевое отклонение и локтевое отклонение. В качестве классификатора была выбрана CNN. Особенностью данного решения является отказ от конструирования признаков в пользу обучения признакам, поскольку из-за природы CNN объекты автоматически изучаются с помощью сверточных слоев. Длина временного окна между предсказаниями была выбрана 285 мс из соображений длительности одного пЭМГ паттерна в 300 мс и времени на его обработку в 15 мс. Таким образом, учитывая частоту дискретизации Myo Armband в 200 Гц на одно предсказание приходится 57 временных точки. Предобработка заключается в использовании преобразования Фурье окном в 30 точек и перекрытием между соседними окнами в 21 точку. Получаемые 4 спектрограммы объединяются в матрицу спектрограмм 16×4, после чего из этой матрицы удаляется первая строчка, содержащая частоты 0 – 6,67 Гц, поскольку этот диапазон находится вне полезного 5 – 450 Гц. Спектрограммы рассчитываются с использованием библиотеки Scipy на языке Python и имеют окончательный диапазон частот от 6.67 до 100 Гц. Все 8 спектрограмм подаются на вход двухэтапной нейронной сети. Если нейросеть обнаружила на первом этапе любой жест, кроме нейтрального, то выходной слой будет иметь 6 нейронов, иначе – 2 нейрона. Использование двухстадийного подхода позволяет значительно улучшить точность классификации, поскольку используемая в нейросетевой архитектуре нормализация увеличивает интенсивность сигнала, соответствующего нейтральному жесту, после чего нейросеть принимает его за другой жест. Нейросеть имеет 8 скрытых слоев, некоторые из которых сверточные, а другие полносвязные (рисунок XXX). В качестве функции активации используется гиперболический тангенс, поскольку он показал себя наилучшим образом в сравнении с ReLU и сигмоидой. Для предотвращения переобучения используется dropout (выбраны коэффициенты 0,25 для верхних сверточных слоев и 0,5, 0,75 для более глубоких полносвязных). Как было упомянуто ранее, перед подачей входных данных они нормализуются (лучший результат показала норма ). Предложенная архитектура показала высокую точность и эффективность. В результате проведенных экспериментов при кратковременной мышечной усталости и длительном использовании точность достигла в среднем 97,9% в течение двух 6-дневных периодов. Было также показано, что система легко адаптируется к различным пользователям. К недостаткам можно отнести большое количество параметров нейросети (3,6 млн), что приводит к длительному процессу обучения и высокой вычислительной сложности обучения и работы нейросети.

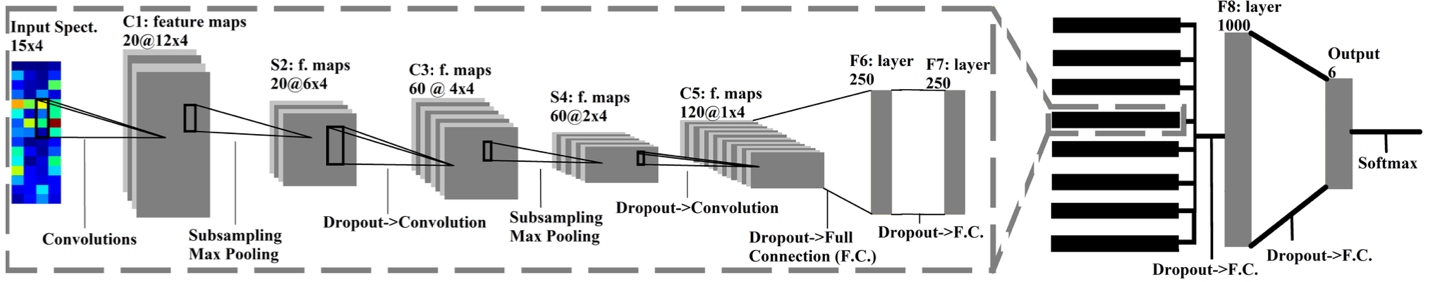


Рисунок 3. Архитектура CNN.

В статье [10] было рассмотрено использование CNN для классификации жестов руки с помощью поверхностной электромиографии высокой плотности (HD sEMG, High-Density Surface EMG). HD sEMG отличается от обычной пЭМГ использованием массива электродов (обычно используется от 32 до 350 штук). Благодаря большому количеству электродов HD sEMG менее чувствительна к расположению электродов, а также более устойчива к неисправностям отдельных каналов регистрации сигнала. Недостатками HD sEMG являются высокая стоимость датчиков и высокая вычислительная мощность, требуемая для обработки сигнала с множества каналов. В качестве обучающих датасетов авторы выбрали открытые базы данных CapgMyo и CSL-HDEMG. Сигналы HD sEMG в CapgMyo регистрировались с помощью электродной сетки размером 8×16, обернутой вокруг правого предплечья. 128-канальные сигналы подвергались полосовой фильтрации на частоте 20-380 Гц и дискретизации на частоте 1000 Гц. Для базы данных CSL-HDEMG сигналы были биполярно записаны электродной сеткой 7×24 (168 каналов). Для поиска сегмента, содержащего активность, к временному ряду интенсивности был применен подход с пороговым значением в скользящем окне длиной 73,2 мс без перекрытия. Значения средней интенсивности в окне усреднялись для всех каналов и по этой величине определялось наличие активности. В связи с тем, сигналы HD-sEMG несут как пространственную, так и временную информацию о мышечной активности, авторы предложили использовать в качестве классификатора CNN с трехмерными сверточными ядрами (3D CNN). CNN с ядрами 2D-свертки ограничены обработкой одного изображения, которое несет только пространственную информацию. Напротив, CNN с ядрами 3D-свертки обрабатывают сразу несколько смежных кадром и потому способны извлекать признаки как из пространственных, так и из временных измерений. Построенная авторами 3D CNN содержит 5 скрытых слоев: 3 сверточных и 2 полносвязных. Опираясь на другие работы, в которых 3D CNN использовалась для классификации видеоданных, авторы выбрали для всех сверточных слоев ядра размером 3×3×3. Более наглядно архитектура нейросети изображена на рисунке XX. Для того, чтобы доказать преимущество 3D CNN над 2D CNN, авторы также построили вторую нейросеть, в которой три трехмерных сверточных слоя были заменены тремя двумерными сверточными слоями. В результате для датасета CapgMyo точность 3D CNN была выше на 1% и 1,5%, чем 2D CNN, для длин окон распознавания 40 мс и 150 мс соответственно. Для CSL-HDEMG точность 3D CNN была на 15,3% и 18,6% выше, чем 2D CNN. Таким образом, авторы статьи делают следующий вывод: для двух CNN с одинаковыми архитектурами трехмерная сетка показывает лучшую точность.

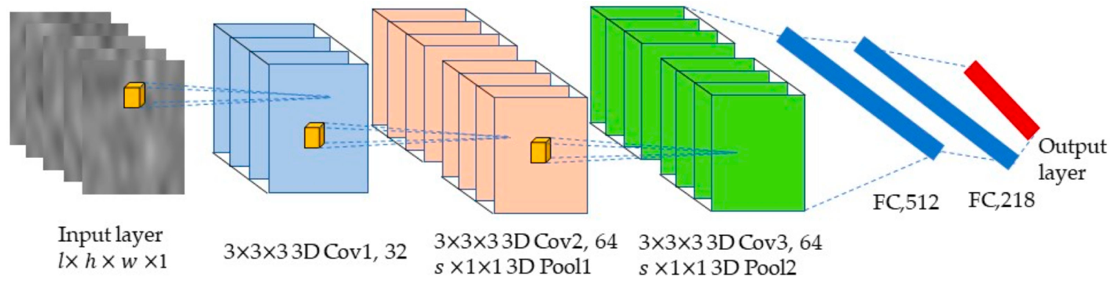


Рисунок 4. Предложенная архитектура 3D CNN.

В статье [11] произведен систематический литературный обзор научных работ, посвященных HGR (распознавание жестов руки) в режиме реального времени с использованием ЭМГ и ML (машинное обучение). По ключевым словам авторы нашли 487 исследований, в результате отсева с помощью анализа названий, тезисов и выводов, а также критериев включения и исключения и вопросов для оценки, были отобраны 65 работ. Авторы рассмотрели 4 исследовательских вопроса:

1. Какова структура моделей HGR в реальном времени, использующих ЭМГ и ML?
2. Какова задержка распознавания и аппаратное обеспечение, используемое моделями HGR реального времени, использующими ЭМГ и ML?
3. Какое количество и тип жестов распознаются моделями HGR в реальном времени, использующими ЭМГ и ML?
4. Какие результаты и показатели используются для оценки моделей HGR в реальном времени, использующих ЭМГ и ML?

В результате комплексного анализа отобранных работ авторы обзора пришли к следующим выводам:

1. Структура моделей варьируется от одной работы к другой, однако можно выделить 6 основных этапов: сбор данных, сегментация, предварительная обработка, извлечение признаков, классификация и постобработка. При этом сегментация, предварительная обработка, извлечение признаков и постобработка являются не обязательными. В 27 работах из 65 для сбора данных использовался Myo Armband с частотой дискретизации 200 Гц, однако, в соответствии с теорией Найквиста и теоремой Котельникова, частота дискретизации должна быть по крайней мере в два раза выше наивысшей частоты полезного сигнала, которая для ЭМГ составляет 400-450 Гц. Для этапа сегментации удалось определить следующую закономерность: при увеличении длины окна, соответствующему жесту, увеличивается точность распознавания, но вместе с тем увеличивается и время распознавания. Наиболее распространенным методом предварительной обработки является фильтрация шума электросети на частоте 50–60 Гц.
2. Задержка распознавания представляет собой сумму двух значений: времени сбора данных и времени анализа данных. При этом можно назвать модель работающей в режиме реального времени, если задержка распознавания меньше некоторого порогового значения, при котором человек вообще не ощущает задержки, называемого оптимальной задержкой. Например, для таких устройств, как клавиша или переключатель, оптимальная задержка составляет 100 мс. Для HGR систем, использующих ЭМГ сигналы, в научной литературе можно встретить различные значения оптимальной задержки от 100 мс до 500 мс для различных приложений. Поскольку используемое аппаратное обеспечение ограничивает вычислительную мощность, оно влияет на время анализа данных, а, значит, необходимо учитывать его влияние при оценке задержки распознавания.
3. Жесты рук делятся на статические (без движения) и динамические (с движением). Сигнал ЭМГ имеет два режима: переходное, возникающее при выполнении динамического жеста или смене жеста, и установившееся, соответствующее постоянному сокращению мышц при удержании жеста. Классификация сигнала в установившемся режиме является более точной, поскольку в этом случае временной ряд является стационарным. По этой причине обычно рассматривают статические жесты, причем важна продолжительность их удержания, поскольку чем она больше, тем больше доля установившегося режима в соответствующем жесту окне. Обычно количество распознаваемых жестов варьируется в пределах от 4 до 20.
4. В соответствии с типами оценки работы были распределены на две группы: модели, использующие метрики ML и модели, использующие тесты достижения целей. В 56 исследований, относящихся к первой группе, использовались 13 оценочных показателей, таких как: доля правильных ответов (accuracy), точность (precision), полнота (recall), доля правильных ответов для одного испытуемого, точность для одного испытуемого, полнота для одного испытуемого и др. Точность – доля правильных ответов среди данных одного класса. Полнота – доля правильных ответов среди данных, распознанных моделью как данные одного класса. Эти показатели могут давать предвзятые результаты по двум причинам: неправильное определение истинного положительного результата и несбалансированный тест. Тестовый набор считается сбалансированным, если в нем присутствует одинаковое количество экземпляров для каждого класса и одинаковое количество экземпляров для каждого участника. Например, если в тестовом наборе 90% примеров принадлежат классу А, то доля правильных ответов всегда предсказывающей класс А модели окажется равна 90%. Важным фактором, влияющим на оценочные показатели модели, является использование обратной связи. Обратная связь позволяет испытуемым корректировать свои движения в соответствии с предсказаниями модели. В большинстве отобранных работ не указано наличие или отсутствие обратной связи. Используя нормальное распределение для количества испытуемых и неравенство Хеффдинга для количества жестов на человека, авторы обзора подсчитали, что размер выборки тестового набора должен составлять порядка нескольких сотен тысяч, чтобы получить результат с уровнем достоверности 95% и пределом погрешности 5%. Ни в одной из проанализированных работ не используется набор тестов такого масштаба, и поэтому авторы обзора приходят к выводу, что достоверность и воспроизводимость их результатов сомнительны.

**??? Общие принципы работы нейросетевых алгоритмов.**

Нейронная сеть (нейросеть) ­– математическая модель, построенная по принципу организации и функционирования биологических нейронных сетей. Простейшей нейросетью является многослойный перцептрон. Рассмотрим многослойный перцептрон, решающий задачу классификации черно-белых изображений (28×28 пикселей) цифр от 0 до 9, взятых, например, из датасета MNIST. В первую очередь необходимо преобразовать файл изображения в некоторый удобный для работы вид. Для этого сопоставим черному пикселю число 0, а белому число 255, при этом все промежуточные полутона серого получат промежуточные значения. В итоге имеем матрицу действительных чисел размером 28×28. Для более наглядного представления преобразуем матрицу к вектору из 28×28=784-мерного пространства. Этот вектор будет являться вектором признаков. Первый слой нейросети является входным, значения на нейронах этого слоя будут тождественно равны значениям входного вектора. Последний слой, являющийся выходным, должен содержать 10 нейронов, при этом задача классификации заключается в том, чтобы значения на этих нейронах отображали вероятность принадлежности конкретной картинки к некоторому классу по мнению нейросети. При этом настоящим ответом задачи классификации будет являться *вектор ответов*, построенный по принципу унитарного кодирования (one hot encoding). Промежуточные слои, называемые скрытыми, могут содержать разное количество нейронов. Передача сигнала с одного слоя на следующий происходит следующим образом: вектор значений на n+1-ом слое есть результат умножения вектора значений на n-ном слое на матрицу, называемой матрицей весов. Таким образом сигнал распространяется от входного слоя к выходному, а значения на выходном слое и есть ответ нейросети. Однако, в таком случае входной вектор каждый раз умножается на i-тую матрицу весов, а, значит, ответом является произведение входного вектора на некоторую итоговую матрицу. Получается, что описанная архитектура является обычной линейной моделью. Чтобы расширить возможности многослойного перцептрона добавим после каждого слоя функцию активации. Функция активации – это математическая функция, применяемая к вектору значений поэлементно и позволяющая усложнить итоговую формулу, описывающую результат действия нейросети. Существует несколько наиболее распространенных функций активации (преимущества и недостатки которых будут описаны далее), например, сигмоида:

*, (1)*

где *x* – действительная переменная.

График сигмоиды, изображенный на рисунке 4, подсказывает, почему функция активации называется именно так: модельные нейроны, подобно биологическим, бывают либо возбуждены при положительном сигнале, либо не возбуждены при отрицательном, а промежуточное состояние достигается только в узкой полосе сигнала.

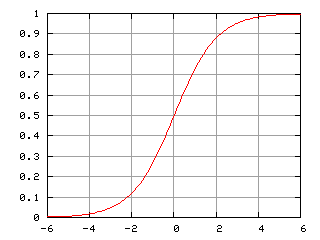


Рисунок 5: График сигмоиды.

Получается, что матрицы весов, связывающие нейроны друг с другом, являются аналогом биологических синапсов, а функция активации описывает нелинейный отклик отдельного нейрона на входной сигнал.

Чтобы описанная выше нейросеть начала работать, необходимо инициализировать все её параметры. Параметрами являются все числовые коэффициенты, полностью описывающие нейросеть (значения весов во всех матрицах, коэффициенты в формулах для функций активации и многие другие). Инициализируем параметры случайным образом. Понятно, что в таком случае нейросеть будет выдавать практически случайные ответы. Чтобы нейросеть решала поставленные задачи (например, всё ту же задачу классификации), необходимо её *обучить*. Классическая схема обучения нейросети выглядит следующим образом. В первую очередь исходный датасет делят на две выборки: обучающую и валидационную. Обучающая выборка используется для нахождения наилучших параметров нейросети, а валидационная для определения точности работы нейросети. Как только нейросеть покажет заранее выбранную достаточную точность работы на валидационной выборке (то есть пройдет валидацию), можно будет сказать, что нейросеть обучена достаточно хорошо. О причине того, почему некорректно использовать для валидации обучающую выборку, будет сказано позже.

Чтобы формализовать решаемую задачу, перевести на понятный нейросети язык, вводят функцию потерь – критерий близости ответов нейросети к истинным (критерий «хорошести» работы нейросети). Чем меньше значение функции потерь, тем ближе выходной вектор к вектору ответов; таким образом, задача обучения нейросети – задача поиска глобального минимума функции потерь в пространстве обучаемых параметров. Нейросети могут содержать сотни тысяч таких параметров, поэтому подбор значений, удовлетворяющих минимуму функции потерь, является сложной задачей и с алгоритмической, и с вычислительной стороны.

Принято минимизировать значение функции потерь итеративно, при этом на каждом шаге рассматривать не всю обучающую выборку, а один батч *–* небольшой в сравнении с размером всей выборки порции. Простейшим методом поиска минимума функции потерь является стохастический градиентный спуск *(SGD).* SGD описывается следующей формулой:

где – значение вектора переменных на следующем (*t+1*-ом) шаге, – на текущем (*t*-ом) шаге, – скалярная функция потерь, – численный коэффициент, отвечающий скорости обучения (learning rate).

Метод называется стохастическим (случайным), поскольку подсчет значений градиента производят не для всей выборки, а для батча, в котором содержатся случайно выбранные наблюдения. Данное решение позволяет колоссально сократить время одной итерации, поскольку резко сокращается количество вычислений, а также количество используемой оперативной памяти (вес всего датасета может превышать объем оперативной памяти вычислительной машины). При этом, происходит усреднение значений градиента по наблюдениям в батче, что уменьшает случайные отклонения от оптимальной траектории спуска. Однако, в настоящее время существуют более продвинутые методы, позволяющие сходиться к минимуму быстрее, а также избегать застреваний в локальных минимумах. Главным недостатком SGD является общий для всех признаков и фиксированный на каждом шаге learning rate (LR). Если LR будет слишком маленьким, то, во-первых, для нахождения глобального минимума потребуется неоправданно большое количество шагов, а, во-вторых, будет велик риск «застрять» в одном из локальных минимумов, поскольку одного шага алгоритма будет недостаточно для «выпрыгивания» из ямы. Использование аналогии с ямой связано с тем, что в случае, когда – достаточно гладкая функция двух переменных, её можно разложить в ряд Тейлора в окрестности точки минимума с точностью до второго порядка; тогда в этой окрестности полученная квадратичная форма задает эллиптический параболоид, визуально похожий на яму. Если LR будет слишком большой, то велика вероятность попадания в бесконечную петлю из прыжков сквозь центр ямы глобального минимума, поскольку каждый шаг будет перебрасывать нас на противоположный «берег» ямы. Первым улучшением SGD является метод SGD with Momentum (стохастический градиентный спуск с импульсом). Идею данного метода проще всего понять на примере движения материальной частицы в двумерном потенциале. Если до этого мы учитывали исключительно форму потенциала и направляли частицу в сторону скорейшего убывания потенциала (геометрический смысл градиента), то сейчас мы учтем текущую накопленную скорость (или, что суть то же самое, импульс) частицы. Алгоритм SGD with Momentum описывается следующими уравнениями:

(2)

где – значение вектора скорости на следующем (*t+1*-ом) шаге, – на текущем (*t*-ом) шаге, – численный коэффициент.

Как было сказано ранее, в отличие от SGD данный алгоритм помимо градиента на текущем шаге дополнительно учитывает градиенты (скорость) на предыдущих шагах. Данное решение позволяет уменьшить влияние неустойчивых и колеблющихся значений градиента, а также «выталкивать» материальную частицу из локальных потенциальных минимумов. Недостатком является то, что подсчёт градиента осуществляется в текущей точке, однако в случае достаточно большого накопленного значения в скорости, получается, что смещение будет в основном определяться именно этим значением, а, значит, значение градиента получается неактуальным для данного шага. Подсчет градиента в вместо оказывается более удачным решением:

(3)

Данный алгоритм носит название SGD with Nesterov Momentum. Более наглядно разница между SGD with Momentum и SGD with Nesterov Momentum проиллюстрирована на рисунке x.

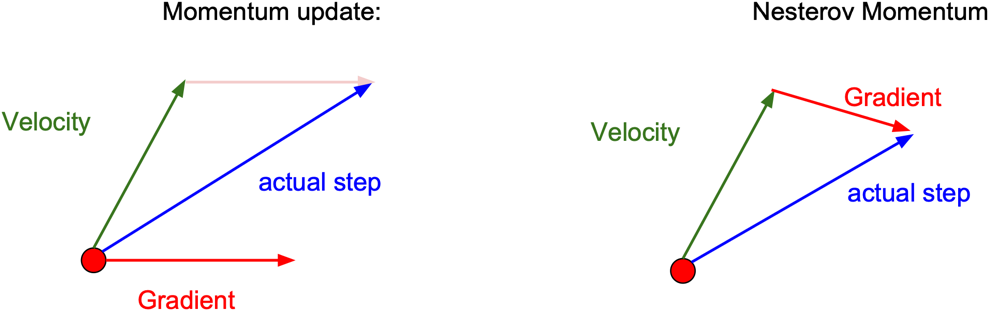


Рисунок 6: иллюстрация различия методов SGD with Momentum (слева) и SGD with Nesterov Momentum (справа).

Существуют и более сложные оптимизаторы. Одним из наиболее широко применяемых оптимизаторов является Adam (adaptive moment estimation), который включает в себя принцип сохранения инерции из SGD with Momentum, но также запоминает и позволяет нивелировать влияние значений градиентов по «шумным» переменным и усиливать наиболее по наиболее значимым.

Любой из перечисленных оптимизаторов требует вычисления градиентов функции потерь по параметрам нейросети. Эта задача также является совсем нетривиальной, поскольку значения на выходном слое нейросети будут зависеть от параметров верхних слоев очень сложным образом. Общим алгоритмом нахождения частных производных по всем параметрам нейросети является метод обратного распространения ошибки (backpropagation). Чтобы понять, почему он называется именно так, рассмотрим пример, когда нейросеть содержит один скрытый слой, тогда схему распространения сигнала можно представить как

(4)

где – входной сигнал, – результат действия первого слоя, – параметры первого слоя, – результат действия второго слоя, – параметры второго слоя. Пусть для подсчёта градиентов необходимо найти . Поскольку зависит от не напрямую, а с помощью , воспользуемся производной сложной функции:

(5)

Таким образом, чтобы найти , сначала необходимо найти , то есть двигаться в обратном обычному направлению распространению сигнала. В случае с настоящими, более сложными нейросетями, поступают аналогичным образом с помощью обобщения данного примера на большее количество скрытых слоёв.

Описанные выше приёмы позволяют решить задачу обучения нейросетей. Однако, оказывается, что сеть можно «слишком хорошо обучить», а точнее переобучить. Феномен переобучения нейросетевых алгоритмов заключается в том, что количество параметров в выбранной модели может быть настолько велико, что позволит нейросети «запомнить» обучающую выборку. Для наглядности рассмотрим в качестве примера регрессионную задачу для n точек на плоскости, имеющих попарно различные координаты по оси OX. Как известно, для такого набора из n точек можно выбрать единственный полином степени не выше n–1 такой, что его график пройдет через все n точек, называемый многочленом Лагранжа. Казалось бы, достаточно найти коэффициенты многочлена по известным формулам, и он будет идеально решать поставленную задачу, ведь его предсказания буквально равны значениям тестовой выборки. Однако, как показано на рисунке x+1, такая регрессионная модель может оказаться абсолютно неподходящей.

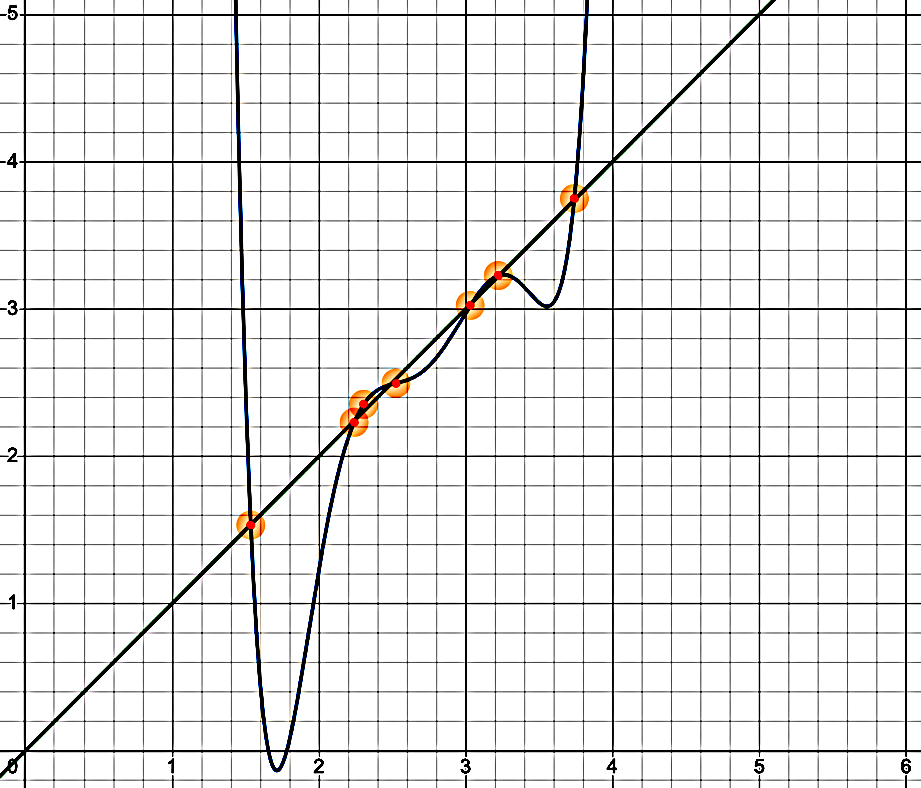


Рис 7. Семь экспериментальных точек, немного не соответствующих зависимости , и их многочлен Лагранжа.

Для избежания переобучения используют различные способы решения данной проблемы. Во-первых, важно разделять датасет на обучающую и валидационную выборки, тогда переобученную нейросеть можно будет легко определить по значительной разнице в точности на этих выборках. Во-вторых, можно применить аугментацию – создание дополнительных выборок для обучения с помощью модификации существующих. В случае решения задачи классификации картинок можно изменять их яркость, резкость, отражать, переворачивать, обрезать и так далее. С помощью аугментации удаётся «принудить» нейросеть извлекать полезные признаки из данных, а не заучивать имеющиеся. Наконец, можно использовать исключение (dropout). Данный способ заключается в исключении или обнулении значений на небольшом количестве случайных нейронах в процессе обучения нейросети. Таким образом можно предотвратить возникновение сложных запоминающих связей между нейронами. Например, нейросеть используется в задаче классификации изображений и «пробрасывает» значение какого-то отдельного признака (цвет пикселя в левом верхнем углу) через все свои слои и решает задачу основываясь только на нём. Dropout разрушает запоминающую связь и принуждает нейросеть к извлечению более сложных и информативных признаков из входных данных.

ДОБАВИТЬ ПРО ДООБУЧЕНИЕ

Рассмотрим подробнее возможные функции активации.

Исследование нейросетевых алгоритмов классификации жестов руки по электромиографическим сигналам

Добавить методы добавить статьи

1 ­– Goodgold J. and Eberstein A. Electrodiagnosis of Neuromuscular Diseases. – 3rd ed. – Baltimore, MD: Williams and Wilkins, 1983.

2 – РАЗРАБОТКА ЭЛЕКТРОДОВ НОВОГО ПОКОЛЕНИЯ

ДЛЯ РЕГИСТРАЦИИ БИОЭЛЕКТРИЧЕСКИХ ПОТЕНЦИАЛОВ СЕРДЦА (<https://cyberleninka.ru/article/n/razrabotka-elektrodov-novogo-pokoleniya-dlya-registratsii-bioelektricheskih-potentsialov-serdtsa>)

3 ­– The extraction of neural strategies from the surface EMG // Dario Farina, Roberto Merletti, and Roger M. Enoka

01 APR 2004 <https://doi.org/10.1152/japplphysiol.01070.2003>

4 – МЕТОДЫ АНАЛИЗА ЭМГ-СИГНАЛОВ: ОБНАРУЖЕНИЕ, ОБРАБОТКА, КЛАССИФИКАЦИЯ И ПРИМЕНЕНИЕ

КОЛОМОЕЦ А.А., КУДРЯВЦЕВ Б.М. // Донской Государственный Технический Университет

5 – Р. Ю. Будко, И. Б. Старченко, Создание классификатора мимических движений на основе анализа электромиограммы, Тр. СПИИРАН, 2016, выпуск 46, 76–89

6 ­– EMG-based Hand Gesture Recognition for Realtime Biosignal Interfacing

Jonghwa Kim, Stephan Mastnik, Elisabeth André

Lehrstuhl für Multimedia Konzepte und ihre Anwendungen Eichleitnerstr. 30, D-86159 Augsburg, Germany [kim@ieee.org](mailto:kim@ieee.org)

7 – Deep Learning for Electromyographic Hand Gesture

Signal Classification Using Transfer Learning

Ulysse Coˆte ́-Allard, Cheikh Latyr Fall, Alexandre Drouin,

Alexandre Campeau-Lecours, Cle ́ment Gosselin, Kyrre Glette, Franc ̧ois Laviolette, and Benoit Gosselin

8 ­– Deep learning for processing electromyographic signals: A taxonomy- based survey

Domenico Buongiorno a,b, Giacomo Donato Cascarano a,b, Irio De Feudis a,b, Antonio Brunetti a,b, Leonarda Carnimeo a, Giovanni Dimauro c, Vitoantonio Bevilacqua a,b,⇑

9 – A convolutional neural network for robotic arm guidance using sEMG based frequency-features

Ulysse Coˆte ́ Allard, Franc ̧ois Nougarou, Cheikh Latyr Fall,

Philippe Gigue`re, Cle ́ment Gosselin, Fellow, IEEE, Franc ̧ois Laviolette and Benoit Gosselin, Member, IEEE

10 – High-Density Surface EMG-Based Gesture Recognition Using a 3D Convolutional Neural Network

Jiangcheng Chen 1,\*, Sheng Bi 1,2,\*, George Zhang 1 and Guangzhong Cao 3

11 ­– Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review

Andrés Jaramillo-Yánez 1,2,\* , Marco E. Benalcázar 1 and Elisa Mena-Maldonado 1