Smart Glove Hand Gesture Recognition: A Literature Survey on Applications and Methods

Khaled Ali Zulfikar
School of Electrical Engineering, Computing and Mathematical Sciences
Curtin University
Perth, Australia
khaled.alizulfikar@student.curtin.edu.au

***Abstract*— Hand gestures, fundamental for human communication and expressing intentions, have gained significant research attention. Noninvasive wearable sensors have emerged as a promising technology for hand gesture recognition (HGR) across various applications including sign language recognition, human-machine interfaces, and rehabilitation. This literature review examines current upper limb wearable devices to investigate feasible solutions capable of clinically accurate HGR. Wearable sensor-based HGR using devices like smart gloves can achieve high accuracy and efficiency compared to vision-based systems. Common sensor modalities include flex sensors, inertial measurement units (IMUs), surface electromyography (sEMG), and capacitive stretch sensors. Machine learning (ML) models like support vector machines and decision trees are commonly used for HGR, while deep learning (DL) algorithms like convolutional neural networks and recurrent neural networks can provide higher accuracy and robustness. Challenges include dealing with variations in hand size, shape, and conditions. The review concludes that smart gloves with a multi-modal sensor design combining flex sensors and IMUs, along with advanced ML/DL algorithms, are a promising direction for clinically accurate HGR. Future research should focus on expanding gesture vocabularies, standardizing sign language detection systems, and developing lightweight, low-power wearable devices.**

Keywords*— *Hand Gesture Recognition (HGR), Human Movement, Sign Language Recognition (SLR), Sensors, Machine Learning (ML), Deep Learning (DL)

I. Introduction

Communication is commonly described as the transmission of information between individuals or groups. It comprises of a sender, a message, and a receiver with the purpose of an understanding [18]. Gestures encompass expressive and concise form of human motion that can involve signals of the hands and fingers, head, face or body and act as a communication technique to be reconstructed by the recipient [17]. Hand gestures are defined as a movement of the hand to express an idea. The objective of Hand Gesture Recognition (HGR) is to classify gestures using measurements into specific gesture classes. Such data would be in the form of a time series, where the specific gesture would occur over a period in that time series [5]. Wearable HGR technology has been widely adopted due to its affordability and reduced safety risks compared to invasive sensing techniques [19]. Recent wearable technology has seen developments in varied forms, including wristbands/arm devices, rings, smartwatches and bracelets. A smart glove, commonly defined as a glove with wearable sensors to extract information based on hand movements [5] is also a method for acquiring signals for HGR. Various wearable sensor modalities can be used to measure environmental and biosignal data, consisting of electrical, mechanical, acoustic and optical methods, such as flexible strain sensors, inertial measurement units (IMUs) and surface electromyography (sEMG). Integration of HGR systems into wearable devices offer solutions to improve the quality of life for speech or hearing-impaired individuals through sign language recognition (SLR) [2,3,4], as well as medical applications [25,36] such as assessing hand rehabilitation progress [25], and human-machine interface (HMI) systems, which recognise human intentions for machine response [2,4,8,35].

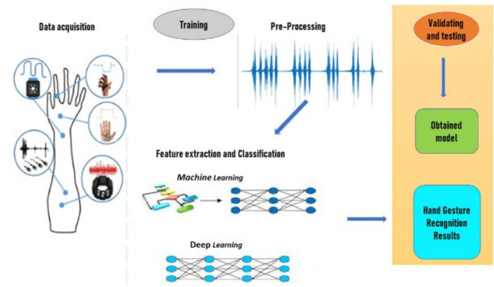


Fig. 1. Data Processing model for Hand Gesture Recognition [19].

Simple lookup table models have been used to decode user gestures from signals produced by wearable sensors. Machine Learning (ML) models, or classification and regression are commonly used in HGR problems to classify gestures from wearable sensor data [37]. ML models require feature extraction techniques, consisting of selecting the best functions that best suit a the nature of the problem. The effectiveness of ML models are influenced by the quality and quantity of data, hence data pre-processing techniques are often used. ML models also tend to achieve higher accuracy when trained on larger datasets. Deep Learning (DL) algorithms are comprised of complex neural networks that can better utilise sensor information without prior knowledge, and can save time and effort as they do not require manual feature extraction, and can automatically discover more complex features than traditional ML models.

A. Objective

The aim of this review is to explore current upper limb wearable devices to investigate feasible solutions capable of clinical accuracy of recognising a predefined set of hand gestures. Signal processing for HGR often follows a structured approach involving: (1) data acquisition (2) pre-processing (3) feature extraction (4) classification. This is shown in [19, Fig. 1]. Designing a smart glove requires careful considerations of each stage, and factors such as hand physiology, sensor placement and algorithm selection must be carefully chosen [5].

The human hand can be modelled as a set of biological sensors and actuators, and the challenge is to determine which sensors can be used to make measurements that are most relevant for gesture recognition [5].

B. Main Contributions and Scope

The review provides a comprehensive survey of available literature for HGR through wearable devices. This literature review excludes developments based on vision-based gesture recognition (VGR) techniques, and the focus is explicitly on upper limb non-invasive sensor-based systems. This literature review aims to evaluate sensor significance and which are most effective considering complexity, costs and prediction accuracy. It is also discusses the electronics circuit design used in smart glove design for power, sampling frequency and computational requirements. Finally, this research will aim to address the difficulties faced in HGR, and address some major algorithms used in recent developments.

This remainder of the review is structured as follows. Section II provides background into different data acquisition mediums, as well as a review on common applications of HGR systems. Section III is an overview of gesture vocabulary requirements. Section IV is a review of recent commonly used sensorisation methods and their electronics interfaces. Section V investigates signal pre-processing methods. Section VI is a review of feature extraction methods for machine learning and sensor fusion strategies Section VII is an overview classification methods such as ML, DL and lookup table/inference models. Finally, section VIII describes future research directions and challenges of HGR systems.

II. Background

A. Data Acquisition Mediums

1) Vision Based

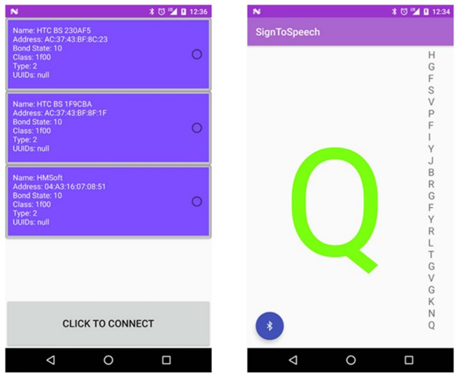
In recent studies, non-contact-based vision-based gesture recognition (VGR) systems classify hand gestures through integrated environmental data acquisition methods, such single cameras, and stereo cameras [9,39]. The sensory data is then analysed through image processing algorithms, leading to the feature extraction and recognition of gestures [39]. Media Pipe Hands and InterHand2.6M are pre-built open source algorithms that operate using a single RGB camera [9]. One common drawback of VGR solutions is the high computational complexity involved in analysing video sequences and are resource intensive [9]. Image based HGR solutions are also restricted to pre-definition contexts, requiring specific lighting and environments such as rooms and apartments. Moreover, it would not be practical for deployment in densely crowded areas due to the interference from other individuals. Comparatively, they tend to be less user-friendly and are not a feasible portable solution required for many HGR applications and might increase users' privacy concerns. Due to this, there is a need for HGR techniques to be employed on a affordable, low powered embedded, or wearable devices with constrained computational capabilities.

2) Wearable Sensor Based

Sensor based gesture recognition technologies (SGR) include wearable devices such as smart gloves, bands and mobile devices [10]. Common SGR techniques implement sensors such as sEMG, IMUs, flex sensors, and photoplethysmography. This data capture method is suitable for embedded devices with limited computational capabilities for capturing and processing sensor data [8].

B. Applications

Extensive research and various approaches have been taken towards HGR, where its applications hold significant importance across various sectors, including consumer, medical, industrial and assistive technology domains. The following sections present the major applications of HGR.



(a) (b)

Fig. 2. Android-based sign language interpretation system showing (a) bluetooth device search screen and (b) text received from wearable smart glove [33].

1) Sign Language Recognition

Non-verbal communication utilising expressions through body movement can be utilised to assist hearing impaired individuals to communicate amongst themselves and with others. Sign languages, such as Australian Sign Language (Auslan), use manual communication through an organised collection of such gestures [9]. These languages are continually evolving over time due to cultural and linguistic norms, and to meet the communication needs of deaf and mute individuals [9]. The problem arises due to the lack of sign language knowledge amongst verbally communicating individuals. This can be solved using a technology-driven solution to facilitate seamless communication and translate gestures into widely spoken languages such as English.

Most research papers focused on SLR have used a limited gesture set from such sign languages, consisting of letters and numbers [3,6,7,33] as well as some employing dynamic word gestures [4,13]. Further progress in SLR can be achieved by expanding the gesture set for a broader vocabulary range, as well as a standardisation of a sign language detection system. In [33], an SLR glove consisting of a sensor module and a processing model were created to wirelessly transmit a recognised alphabet to an mobile device via Bluetooth. An Android-based mobile application on the device was designed with a text-to-speech feature for an audible voice output as shown in [33, Fig. 3].

2) Human Machine Interface

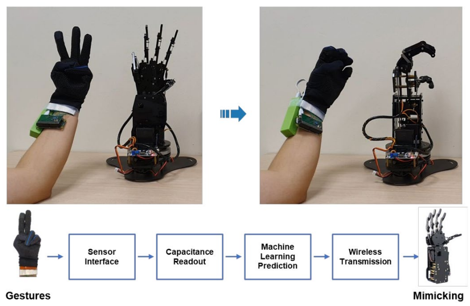


Fig. 3. Schematic showing HMI implementation with mechanical hand for gesture mimicking [2].

HGR can facilitate intelligent communication amongst humans and computers through HMI systems. A common application is an alternative to traditional input devices such as keyboard and mouse [8], where gestures serve as commands for various operations. This provides alternative input methods for immersive virtual reality and augmented reality experience and gaming. Another application is robot control, hazardous environments, or where conventional communication methods are redundant. In [34], CaptAinGlove, used inertial and capacitive sensing methods and a DL algorithm for real-time HGR for drone control. It defined 8 hand gestures for the drone control dictionary. Similarly, a recent study [2] created a robotic mechanical hand consisting of servo motors, a Bluetooth module and a microcontroller. A lookup table of Pulse Width Modulation (PWM) signals were sent to the servo motors were used to replicate a corresponding recognised gesture performed on a smart glove on the mechanical hand remotely, as showing in [2, Fig. 3]. Another study [35] uses a situation-traffic VGR system for directing autonomous vehicles through a traffic controller. It uses a gesture set based on commonly recognised traffic control hand gestures.

3) Rehabilitation

HGR using wearable sensor technology can play a crucial role in motor function recovery following conditions such as stroke as well as diseases and disorders [19].

A recent study [36] is aimed to address rehabilitation challenges of patients with paresis. The HGR glove was created to recognise gestures on the non-affected hand during rehabilitation training using sEMG sensors as well as a DL algorithm. A motorised glove is then worn on the affected hand to provide assistive force using linear integrated actuators on each individual finger. Based on paresis rehabilitation methods such as mirror therapy and task oriented therapy, the hand gestures performed on the HGR glove are then used to control the motorised glove to mimic gesture movements.

In a recent study, non-invasive sensors and machine learning algorithms are employed to precisely measure and record joint movements, offering real-time feedback on exercise progress, as shown in [25, fig. 4]. This framework utilises a game based hand rehabilitation system where wrist movement activities such wrist flexion/extension, ulnar/radial deviation, pronation/supination and finger thumb touching exercises are given to patients for active hand function training. The gamification aspect of these systems enhance patient engagement, which is crucial for long-term rehabilitation success. Wrist exercises were performed using IMU data, whilst VGR techniques were deployed for finger touching exercises. Sensorised HGR can be adopted as an alternative input method, allowing for a deeper level of movement information to be captured.

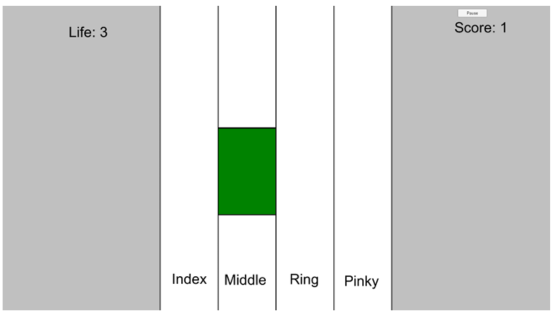


Fig. 4. Gamification of thumb touching exercise for remote hand function rehabilitation training [25].

C. *Anatomy and Range of Motion of the Human Hand*

The human hand is comprised of the carpus (palm) region, the metacarpus (in between the carpus and the knuckles) and the fingers. The four fingers are divided into three phalanges (segments) and the thumb contains two phalanges [39]. Starting from the fingertip, these are described as the distal (proximali), middle (medialis), and proximal (proximalis) phalanx [39]. The two hinge joints separating the phalanx on the four fingers are called distal (DIP) and proximal (PIP) interphalangeal joints respectively [39]. The saddle joint connecting the metacarpals to the phalanges are called metacarpal-phalangeal (MCP) joints. The thumb consists of a distal and proximal phalanx, a single hinge MCP joint as well as a spheroidal ball (thumb-carpo-metacarpal, T-CMC) joint connected to the carpals with three degrees of freedom [39]. The anatomy model of human hand movement can be classified into abduction/adduction finger movement, ulnar/radial deviation of the wrist, wrist extension and wrist flexion as well as finger flexion and finger extension [39].

III. Gesture Vocabulary

Applications such as SLR necessitate both static and dynamic hand positions and movements as shown in [13, fig. 6]. In American Sign Language, most letters are represented by static poses without any accompanying motion [13]. Recent smart glove developments [13] use a gesture set consisting of various letters, words and phrases obtained from American Sign Language. It was found some gestures shared the same static pose but exhibited different dynamics, as well as some gestures exhibiting periodic motions, with variations in repetition. A gesture vocabulary must be chosen to evaluate the systems capability to integrate both pose and motion information for detecting multiple classes of gestures in real time.

IV. Sensorisation and Electronics Design

Movements performed during hand gestures can be biologically detected by muscle activity, tendon movement, blood vessel distortion and bone movement [19], and the most accurate data collection method must be based on these changing biological and physical characteristics relating to the anatomy of the human hand [1]. In order to capture the gesture, sensing modalities can be generalised into four categories: electrical, mechanical, acoustic and optical. Recent studies have shown these different sensor modalities being incorporated on wristbands and smart gloves to detect hand motion data.

It can also be useful to analyse the electronic circuit used for each sensor in previous designs, for analysis of power and sampling frequency limitations, as well as factors such as noise reduction. According to [20], the sampling frequency

Fig. 5. Example gesture set derived from American Sign Language consisting of static poses and dynamic gestures [13].

must be large enough to capture signal variations without losing information. The study indicates that all body movements typically fall within the range of frequencies from 0 to 20Hz. As per the Nyquist criterion, the sampling frequency should exceed twice the highest frequency present in the signal being sampled [19], hence the sensor sampling frequency must be over 40Hz, as this is the minimum frequency that will accurately capture human movements with no loss of information.

A. Mechanical Sensing

The human hand can be modelled as actuators that operate within 25 degrees of freedom (DoF) [1]. The most straightforward solution is to implement a sensor that will account for every DoF, however, not all the information will be of significance. As it is difficult to predict the ranking of importance of each mechanical sensor, this section of the literature review aims to compare and evaluate the performance commonly used glove sensors such that it is to operation within requirements. Recent studies have shown the potential of forcemyography (FMG), inertial measurement units (IMUs) and strain sensing for HGR.

1) Forcemyography

FMGs measure force variations due to muscle movements, including volumetric and stiffness changes and can utilise sensor types such as piezoresistive, pneumatic-based, piezoelectric, force-sensitive resistors (FSRs) [19, 22], and capacitive [2]. In [22], an armband created using six force sensing resistors was used to decode various gestures using ML algorithms, and it showed high potential for recognising pinch, power, tripod and extension gestures when compared to EMG sensing.

a) Bimodal Capacitive Force/Proximity Sensor

Recent research has shown that a custom designed bimodal capacitive sensor array can capture a higher resolution of information about the user's hand movements, including both force interactions involved with flexing, clenching, and grasping up to 30N, and proximity data up to 16cm [2]. A silver-MWCNT composite is then aerosol jet printed directly onto the Kapton sheets, combining the conductivity of the silver and the robustness of MWCNT [2]. This glove comprised of an array of 16 bimodal capacitive sensors, having a size of 5mm x 5mm each and positioned at the fingertips (contact points) and hand joints to maximise sensing resolution. The capacitance of the sensor decreases upon finger approach, and upon contact increases linearly based on force applied, and data algorithms are used to separate the two sensing modes at the backend [2]. The glove also used a CMDA power chip, and combined with passive capacitive sensors had an extremely low power consumption at 8.2µW at 2kHZ.

2) Strain Sensors

Strain sensors, known for their affordability, lightweight build, and stretchable nature, are commonly used to detect the elastic deformation of skin, muscles and tendons [23].

a) Flexible Bend Sensors

Commercially available flexible bend sensors can be used to measure strain by measuring the angle of bending [19]. They can be modelled as variable resistors, where its total resistance is proportional on its radius of curvature, or the amount of bending [7, 12]. Flex sensors typically exhibit resistance values ranging from 7 to 26 kilohms (kΩ) [8]. Flex sensors can be incorporated through a voltage divider circuit, with its output then digitized by an Analog-to-Digital Converter (ADC) integrated within a microcontroller [8]. The output voltage across the flex sensor can be captured using a microcontroller, such that:

$$V_{out} = V_{in} \frac{R_o}{R_o + R_{flex}} \quad (1)$$

Where V_{in} Denotes the supply voltage, V_o denotes the sensory data and R_g indicates a fixed resistance used in the voltage divider circuit. A parameter, B, can be defined as the ratio of V_o to V_s such that the flex sensors resistance R_{flex} is independent of variations in the supply voltage.

$$B = \frac{V_o}{V_{in}} \quad (2)$$

R_f can then be derived from eq. (1) by substituting (2).

$$R_f = \frac{(1-B)R_g}{B} \quad (3)$$

Recent developments [6,8] use single flex sensor for each finger. Chuang's model [8] incorporates a LilyPad Arduino as the microcontroller for data collection where five 3.5" flex sensors are connected to available ADC ports. Powered by a LiPo rechargeable battery with a capacity of 70mAh, the module transmits data via an integrated wireless module supporting Bluetooth 4.0. This glove presents benefits including lightweight construction, high power density, and an extensive number of charge cycles. The glove also incorporates e-textile techniques to integrate and connect the components within the glove which ensures flexibility and comfort. Chuang's glove was designed to only be responsible for data acquisition, and the HGR process would occur externally through the real-time transmitted data via the Bluetooth module at 9600 bps.

Flex sensors are highly flexible, portable and can be used to accurately capture hand joint information. They must, however, be used in conjunction with other sensors as they cannot capture hand motion data and they can be prone to damage and have repeatability issues [14]. Their accuracy can also be compromised if the sensors midpoint is not precisely aligned with the PIP joint [19].

b) Resistive Knit Glove as Strain Sensor

Recent studies [4] have shown that a commercially available knitted conductive glove, Original Sport glove by Angloves has significant sensitivity to strain. This glove is utilised by weaving approximately 3cm of stripped wire through the knit, and the resistance between the connections serves as a strain sensor. The article describes the measurement of resistance through 16 readout points and a shared ground embedded within the sensor glove. These 16 points were strategically placed between each finger segment for maximum information. This glove has a variable resistance of around 5Ω/m and is suitable for detecting deformations due to the bending of the finger joints. The glove also leads to significant resistance changes when parts of the hand are touching. It was found that the average response time of this glove when explored with gesture detection was around 0.35s, and a recovery time of 0.8s, and can withstand strain without permanent deformation, indicating high repeatability. In the article, the glove was also embedded with a 9-axis IMU.

The data acquisition system is designed for low noise resistive sensor readings that can be embedded in a compact and wearable form. A STM32H7 microcontroller was used for real time data acquisition using the ADC as well as signal processing and machine learning. Two DACs controlled the analogue front end for current regulation and offset removal. The 16 strain sensor readouts were multiplexed via two 16 channel switches, which were sequentially polled. This allowed the specification of the current and offset voltage specific to each sensor. The single IMU was connected via I2C. The microcontroller also contained a Bluetooth transmission for optional wireless transmission, and current tests can be measured through a USB connection.

c) Capacitive Stretch Sensor

In [14], five capacitive stretch sensors were employed to track the bending state of the fingers, as well as a six-axis IMU to capture spatial motion, as well as an environmental sensor to detect water for amphibious HGR. An Esp32-S3-DevKitC-1 board was used alongside a Wi-Fi and Bluetooth interface module.

3) Inertial Measurement Units

Advancement in semiconductor technologies have adopted IMUs, which have been used in some studies to track finger joint movement [15], spine posture and hand movement [6] 3-D space. Typically, smart gloves are equipped with a network of 6-axis or 9-axis IMU sensors, each consisting of accelerometers, gyroscopes and magnetometers which provide output for acceleration, angular velocity, and magnetic field measurements, respectively [16]. By integrating the accelerometer and gyroscope data over time, the IMU can determine its position and orientation. Additionally, the magnetometer enables the IMU to establish its relative orientation to the Earth's magnetic field. [6].

IMU sensors can be used to capture spatial information on a smart glove and are suitable for recognising large gesture movements [6,14]. IMUs can also provide information on overall movement characteristics, such as smoothness and velocity [19]. Recent research has shown that an IMU can be placed on each phalanx for accurate finger joint calculation [15]. Integration over time cause drift issues in the calculated data, and hence Filtering such as Kalman and Complimentary can be used for data fusion and the minimisation of error [6]. Despite the capability of IMUs, they are rarely used on their own, as kinematic and orientation information alone are inadequate for detecting hand gesture patterns [19].

B. Electrical Sensing

Electrical sensing technologies like electromyography (EMG) [21] or electrical impedance tomography (EIT) [19] offer methods for measuring muscle contraction. The difference between the two sensing methods is that EMG sensors detect electromyographic signals produced by muscles during contraction, whilst EIT sensors measure the response to an externally applied electrical current [19]. In reference [23], a wrist-worn system comprising an EMG sensor operating at a sampling rate of 2000Hz was developed and validated for HCI, achieving an accuracy exceeding 95%. Whilst these signals can be used to provide detailed information on muscle activity, they are prone to instability due to the user's physiological state and cannot be adequately integrated in a smart glove as they may be uncomfortable and are affected by oily skin and sweat [40].

C. Acoustic/Vibratory Sensing

Changes in the physical structure of the wrist and forearm can lead to variations in acoustic properties, which can be identified by acoustic sensors in hand gesture recognition (HGR) systems [19]. Acoustic sensors commonly employed in HGR include sonomyography (SMG), mechanomyography (MMG), and bone-conducted sound sensing [19]. Sonomyography (SMG) utilizes ultrasound imaging to gather insights into the deeper tissue layers of muscles and tendons. Recently, it has been utilized to enhance finger motion recognition at a deeper level compared to electromyography (EMG) [24]. SMG is based on an ultrasound signal propagating through the human body, and where the signal will be reflected when it encounters a tissue mutation interference [24]. When used on the forearm, muscle deformation under different gestures can be detected and decoded by grouping the echo signal of muscles into gesture groups [24].

D. Optical Sensing

Recent advancements have seen optical sensors being integrated into smartwatches, such as Photoplethysmography (PPG), commonly used as a lightweight, inexpensive heartrate monitor [19, 27]. PPG sensors are comprised of a light source and a photodetector, that can measure volumetric variations of circulating blood through the reflection of the skin [27]. A recent study [28] has shown that an accuracy of 98% can be achieved when differentiating nine gestures from the American Sign Language (ASL) using a combination of PPG and motion sensors.

E. Multimodal Sensing

Recent studies have shown that the integration of data from multiple sensors [5,6,14] through multi-sensor fusion techniques can leverage the strength of different approaches and capture different aspects of a gesture, improving overall gesture recognition accuracy, and capture a wider range of gestures, but with a downside of complexity and cost.

1) Sensor Importance Estimate Glove

A recent study [5] was structured to accommodate as many sensors as feasible, even if it resulted in redundant information. It had a total of 28 sensors including five pressure sensors on each fingertip, 13 flex sensors on fingers and wrist, seven IMUs on the wrist and fingertips and a single magnetometer as shown in [5, fig. 7]. The sensor information is captured using an Arduino Duo board, where the IMU and magnetometer are connected via an I2C bus, and analogue sensors including flex and pressure are connected using a multiplexer and read every 12mS (83.3Hz). The sensors are

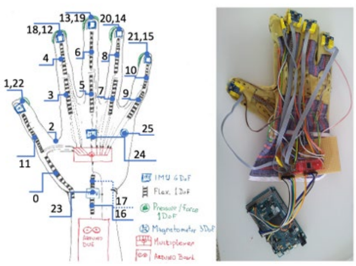


Fig. 6. Importance estimate glove consisting of IMUs, flex sensors, force sensing resistors, magnetometers, a multiplexer and Arduino board [5].

then initialised and captured using a PC via a serial connection.

V. Data Pre-Processing

Data pre-processing involves the removal of unwanted features from a signal such as electrical or ambient noise to ensure optimal data quality, as they may affect HGR accuracy [19]. Recent studies are using IMU and FMG sensors without filtering as they have acceptable Signal to Noise Ratios [30]. Other filtering techniques such as pre-smoothing, offset compensation, amplification, and rectification have also been used depending on sensor type and quality [19]. Sampling techniques are additionally employed to prepare the signal for the subsequent classification stage, facilitating more efficient data collection. Variable-rate, fixed-rate, adaptive, sensor-bit resolution tuning, and compressed sensing are sensor sampling techniques aimed at optimizing data collection through various methods [19]. Time synchronisation techniques such as dynamic time warping (DTW) are also deployed as a solution to synchronise multimodal inputs, preventing mismatched data in time leading to incorrect predictions [19].

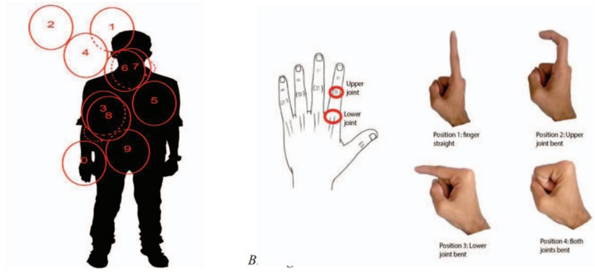
VI. Feature Extraction

Machine learning models are linked to the challenge of dimensionality. When dealing with a large number of features relative to the amount of data available, overfitting becomes a concern, making training difficult in practice [37]. Hence, dimensionality reduction techniques, or feature extraction methods are applied. Feature extraction entails converting raw data into a set of features better suited for classification, hence the target is to select a subset of input features from a dataset [19, 32]. DL models derive high-level features automatically through representation learning, while ML models will require manual feature selection [19]. In ML, features can be extracted from digital signals through (1) the time domain (TD) (2) the frequency domain (3) the time-frequency domain (TFD) [19]. Commonly used features are statistical indexes, such as root mean square (RMS), mean, max, standard deviation, skewness, kurtosis, variance, zero crossing etc. Recent studies extract many features from raw data, and use a feature selection tools to decide which features are most effective [32].

Feature selection methods can be either supervised or unsupervised [32]. Unsupervised algorithms do not use the target variable, or labelled output data, but aim to find patterns, correlations and structures in the data without explicit guidance, and high correlations between variables indicate redundancy [32]. Supervised algorithms learn from labelled output data, where variables that don't contribute to predicting a target variable are deemed redundant and removed [32]. Supervised methods are divided into wrapper, filter and intrinsic as shown in [32, fig. 8]. Supervised methods are more commonly used for HGR problems.

A. Sensor Fusion Strategies

A recent study [31] has used a combination of sEMG, and IMU signals with an objective of designing an accurate neural network comparing different sensor fusion strategies: data-level fusion, feature-level fusion, and decision-level fusion. Data-level fusion consists of concatenating the three sensor input vectors into one, and directly feeding it into the deep belief net (DBN) neural network. In contrast, feature level fusion consisted of separately training DBNs for each sensor input to learn features, extracting learned features from each DBN, and concatenating them into a single input vector for another DBN. Decision level fusion starts by training separate DBNs for each sensor input, extracting soft-max layer outputs from each DBN, averaging these outputs, and using the average to determine the final classification result.



(a) (b)

Fig. 7. (a) States areas for state estimation algorithm identifying position of hand in 3-dimentional space. (b) Discrete voltage ranges are set based on the position of the fingers in the figure [6].

Using a gesture set comprising of 150 sub words from Chinese Sign Language (CSL), it was found that feature level fusion produced the most accurate results, at 95.1%.

VII. Classification

The main objective of classification models is to recognise a set of hand gestures in real-time, without capturing unintended gestures, which requires for the model to have an accurate understanding of gestures. ML and DL wearable systems are able to recognise gestures based on learning from past experience, and hence eliminate the need for explicit programming [19].

A. Lookup-table and Inference Methods

Many existing SGR systems have the shortcoming of recognising gestures by means of a lookup table and inference models [6]. This method is inadequate, as they may not generalise well to new users, environments, and conditions, as well as lacking adaptability to unseen gestures. They also have limited scalability and would be difficult to implement for a complex gesture set and are an inefficient use of memory, which leads to an increase in processing time and power usage [2]. Continuous gesture recognition, involving the classification of a sequence of gestures also poses a challenge with these methods [8]. This challenge involves spotting of the start and ending positions of each gesture in a sequence so each gesture can be isolated and independently recognised. These SGR systems will likely require user program or user assisted gesture spotting methods.

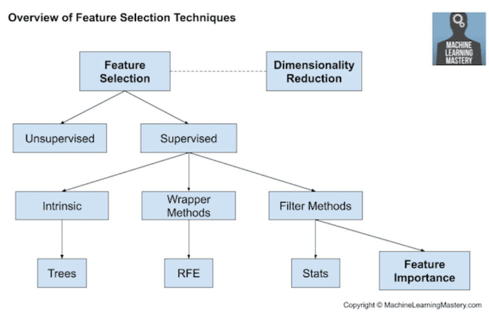


Fig. 8. Overview of feature extraction techniques for Machine Learning [32].

1) Finger Position and State Estimation Algorithm for Dynamic HGR

To detect finger orientation, sensor data from finger tracking methods such as flex sensors can be divided into discrete voltage ranges. Bhaskaran's glove uses a and classifies four discrete voltage ranges dependant on the amount of bending of the combined finger joints as seen in [6, fig. 9]. The glove also uses a 'State Estimation algorithm', using the accelerometer values and gyroscope values to identify the position of the hand in three-dimensional space [6]. Using these values, the hand position relative to the body can be split into a pre-defined discrete set of states, each bounded by an area of variable size. The algorithm is then able to classify whether the gesture is static or dynamic depending on whether the hand is in the same position for a given amount of time. The program then uses values from the IMU to capture state patters (such as 4,4,2,2 for gesture Bye), and a corresponding finger position that maps to a gesture over a given amount of time. This allows for any combination of static and dynamic gestures.

B. Artificial Intelligence Methods

Sensor-based HGR algorithms can be realised through traditional machine learning methods (ML), as well as deep learning (DL) approaches. The choice between ML and DL depends on problem nature, available data and available computational resources [19]. ML-based HGR have lower computational requirements compared to DL methods. ML algorithms such as Dynamic Time Warping (DTW) and support vector machines (SVM) are versatile and can be deployed for various gesture data types [2,14]. However, as HGR tasks become more complex, they will require larger training datasets, which will impact accuracy and processing time [14]. DL algorithms such as (CNNs), (DNNs), and Recurrent Neural Networks (RNNs) have been implemented in various smart gloves, and offer superior, efficient performance due to their ability to handle sequential data [14]. Recent developments have used DL algorithms for SLR and complex dynamic HGR, at high accuracy [2,5,8,13]. DL algorithms can be adopted to provide better performance and efficiency for complex HGR. In recent studies, ML techniques have been commonly used in recognising simple static gesture sets, whilst complex, dynamic gestures heavily rely on DL techniques.

1) Training Data

In a HGR, supervised training methods are favoured, where data from each gesture will be discretely recorded and labelled [19]. Training data should be aimed to explore classifier robustness across episodes of use [13], and consider various factors such as glove positioning variations, wear and tear, hand conditions (temperature, moisture etc.), size of hand, hand shape and other factors. A large time series sample data set, containing many participants performing various gestures will increase the accuracy of the gesture prediction model [5]. In [5], data of 31 gestures were collected from a total of 22 participants. The participants were shown how to perform a specific gesture, and then they would perform the gesture 5-10 times whilst being recorded.

2) Machine Learning

ML models can be categorised into supervised, unsupervised and reinforcement learning [19].

Supervised learning relies on labelled training data for classification and regression tasks. These can include support vector machine (SVM), Decision Trees (DT), and random forest (RF) classifiers [19]. In [2], various supervised ML algorithms were tested including Naive Bayes (NB), logistic regression (LG), RF, extremely randomized forest (XRF), quadratic discriminant analysis (QDA), and multilayer perceptron (MLP). A QDA model was deemed to be the most favourable, as it scored second most accurate at 99.69% and 99.56% respectively to an unseen test dataset and cross validation data set of 467,236 samples. Moreover, its fast prediction time of 10μS per sample and low memory requirements of 85kB deemed the method most suitable for important time and power requirements. In [14], it was found that recognising precise dynamic gestures underwater was challenging due to external disturbances such as water pressure, resistance and flow that would disrupt IMU signals. Hence, it was proposed that a model would be used to switch between static only and static and dynamic gesture sets when underwater and on land respectively by using a barometric sensor for environmental recognition. In this research paper, a lightweight stochastic singular value decomposition (SVD) optimised ML model was used to recognise static gestures relying only on stretch sensor data when the glove sensed it was underwater. This model efficiently clusters data by mapping high-dimensional matrices to low-dimensional subspaces whilst preserving original matrix information [14]. The model also used a SqueezeNet-BiLSTM algorithm for complex dynamic gestures, discussed in 3).

A challenge with supervised ML algorithms, however, is that they can be time-consuming and costly due to the requirement of data labelling, and can be challenging when multimodal sensing is used. Unsupervised machine learning (ML) models, such as Latent Dirichlet Allocation (LDA), Principal Component Analysis (PCA), and Self-Organizing Feature Maps (SOFMs), are employed to extract generative features and uncover meaningful trends and structures. In [2], a dataset comprising over 2.3 million samples was collected from three adults executing various static gestures in different orientations across multiple trials using a wearable device. The gesture set encompassed a total of 26 letters and 10 numbers. LDA was applied to the dataset to reduce data dimensionality and visualize gesture distinguishability. The analysis revealed that the gestures had been segregated into clusters, with sub-cluster differentiations observed, which is advantageous for accurate classification.

Reinforcement learning applies rewards for desired behaviours and punishments for undesired ones, however, this method is less common in HGR [19].

As no classification model can suit all applications and datasets, recent studies experiment with various algorithms and choose based on requirements such as accuracy and computational complexity.

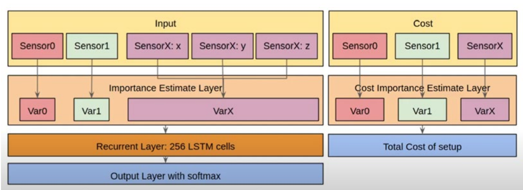


Fig. 9. Overview of importance estimation HGR model accounting for sensor importance and cost [5].

3) Deep Learning

DL methods can present many benefits in the case of HGR, including high accuracy and robustness, automatic learning of deeper features from raw data without manual feature extraction, as well as the ability to extract cross-modal features when multimodal sensing is used [19]. DL algorithms including Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANNs) can automate the feature extraction process with the ability to classify unstructured data. DL methods may not always guarantee better performance than ML, as they may require substantial data inputs in order to achieve high accuracy [19], hence considerations such as dataset size, task complexity and computational requirements should be taken into account.

A Long-short Term Memory (LTSM) neural network, a variant of a recurrent neural network (RNN) is a mathematical model that can take a time series sample as an input and output a probability distribution of which the gesture the sample belongs to [5]. Research has shown the models potential to recognise complex dynamic gestures using sensor data due to its ability to utilise temporal memory characteristics from full sequences of data and overfitting prevention properties [2, 5, 13]. In [2], an LTSM network was trained using the proximity information of the 16-channel capacitive sensor dataset. Three volunteers were used to performed three dynamic gestures in succession. It was found that each gesture contained unique patterns and causing similar fluctuations in the signal. The model was tested using a 500-sample input window, and achieved over 99.7% accuracy on gestures performed individually, and 91.6% accuracy with a sample including unseen gestures.

In [5], an LTSM network and information from a baseline gesture recognition model was used to analyse the importance of each sensor used in the study. From the model, sensor importance estimates were generated by comparing the significance of sensors in relation to the classification of gestures. This was observed by introducing an importance estimate layer between the input and the neural network as seen in [5, fig. 10]. A cost importance estimate layer was also added to account for the additional constraint. The study was focused on obtaining a personalised glove solution through the removal of redundant sensors for individual participants based on the importance estimate model. It was found that the flex sensors on the base and tip of the thumb, pressure sensors in the middle and ring finger as well as IMU's on the wrist were commonly removed between participants.

In [14], a SqueezeNet-BiLSTM model was adopted to extract deep features and time series features for HGR. This model is a combination of a SqeezeNet network and a Bi-directional LSTM network. The collected gesture information is firstly pre-processed through steps of scaling to uniform length, sliding window, filter processing, standardization, normalization, and Tucker decomposition. The processed data is then entered into the SqeezeNet network in order to obtain feature vectors by reducing model parameters through convolution modules, maximum pooling layers and fire modules. A bi-directional LSTM network was then used as a gesture classifier, which addresses LSTM's limitation of only predicting the next moment's output based on a previous moments timing information. This model uses the combination of two LSTM layers to utilise forward and backward feature information from an input gesture data to provide better context for the output.

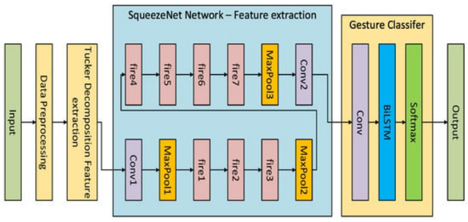


Fig. 10. Overview of SqueezeNet bi-directional LSTM neural network model for HGR [14].

A Gated Recurrent Unit (GRU), [11] is similar to LSTM and is a variant of recurrent neural networks (RNNs), commonly employed for sequence modelling tasks. It effectively captures temporal dependencies within sensory data while boasting lower computational complexity compared to LSTM [8]. The GRU takes an input sequence $X = \{x_1, ..., x_T\}$, where T signifies the sequence length, and generates a corresponding state sequence $H = \{h_1, ..., h_T\}$, each state h_i having dimension D. When initialized with $h_0 = 0$, H can be computed from X. Chuag's model [8] utilizes sensory data obtained from flex sensors on a wireless glove in the sequence $S = \{s_1, ..., s_n\}$, where s_i represents a sample acquired at time step t. The GRU algorithm is employed for gesture spotting operations, yielding gesture spotting results $Y = \{y_1, ..., y_n\}$. All samples in S are vectors with a constant dimension D, determined by the type of sensors utilized for HGR. During the training process, the model considered both finger movements and transitions between successive gestures. The model adopts a labelling scheme for the training data, where each finger gesture and transition share the same label, ensuring that transitions are incorporated into the training process.

VIII. Conclusions

In this study, recent advancements in smart glove HGR technology were examined and their practical applications, with a focus on glove sensorisation and electronics design as well as data processing and feature extraction methods through ML and DL. The numerous studies aimed at comparing the diverse detection modalities, proposing algorithms for effective gesture classification and detailing the innovative development of wearable sensors for sensory data capture. Regardless of the sensing methods, DL and ML integration into wearable systems are covered in the review and are becoming increasingly feasible techniques. Overall, the study suggests a significant potential for HGR to be applied for industrial and clinical settings, with a high potential in application areas such as SLR, rehabilitation and HMI domains. Moving forward, research efforts should prioritize broadening gesture vocabularies, establishing standardized sign language detection systems, and creating lightweight, energy-efficient wearable devices.

References

[1] J. Chan, E. Veas, and J. Simon, "Designing a Sensor Glove Using Deep Learning," Apr. 2021, doi: https://doi.org/10.1145/3397481.3450665.

[2] J. Pan et al., "Hybrid-Flexible Bimodal Sensing Wearable Glove System for Complex Hand Gesture Recognition," ACS Sensors, vol. 6, no. 11, pp. 4156–4166, Nov. 2021, doi: https://doi.org/10.1021/acssensors.1c01698.

[3] K. S. Abhishek, L. C. F. Qubeley, and D. Ho, "Glove-based hand gesture recognition sign language translator using capacitive touch sensor," IEEE Xplore, Aug. 01, 2016. https://ieeexplore.ieee.org/document/7785276 (accessed Jul. 02, 2022).

[4] J. DelPreto, J. Hughes, M. D'Aria, M. de Fazio, and D. Rus, "A Wearable Smart Glove and Its Application of Pose and Gesture Detection to Sign Language Classification," IEEE Robotics and Automation Letters, vol. 7, no. 4, pp. 10589–10596, Oct. 2022, doi: https://doi.org/10.1109/lra.2022.3191232.

[5] J. Chan, E. Veas, and J. Simon, "Designing a Sensor Glove Using Deep Learning," Apr. 2021, doi: https://doi.org/10.1145/3397481.3450665.

[6] K. A. Bhaskaran, A. G. Nair, K. D. Ram, K. Ananthanarayanan, and H. R. Nandi Vardhan, "Smart gloves for hand gesture recognition: Sign language to speech conversion system," IEEE Xplore, Dec. 01, 2016. https://ieeexplore.ieee.org/abstract/document/7931887 (accessed May 03, 2023).

[7] D. Konstantinidis, K. Dimitropoulos, and P. Daras, "SIGN LANGUAGE RECOGNITION BASED ON HAND AND BODY SKELETAL DATA," IEEE Xplore, Jun. 01, 2018. https://ieeexplore.ieee.org/document/8478467 (accessed Apr. 25, 2023).

[8] W.-T. Chuang, W.-J. Hwang, T.-M. Tai, D. Huang, and Yun-Jie Jhang, "Continuous Finger Gesture Recognition Based on Flex Sensors," vol. 19, no. 18, pp. 3986–3986, Sep. 2019, doi: https://doi.org/10.3390/s19183986.

[9] S. Khaksar, L. Checker, B. Borazjan, I. Murray, "Design and Evaluation of an Alternative Control for a Quad-Rotor Drone Using Hand-Gesture Recognition," Sensors, vol. 23, p. 5462, 2023. [Online]. Available: https://doi.org/10.3390/s23125462

[10] H. K. Sharma and T. Choudhury, "Applications of Hand Gesture Recognition," in L. Kane, B. Dewangan, and T. Choudhury (Eds.), Challenges and Applications for Hand Gesture Recognition, pp. 194-207, IGI Global, 2022. [Online]. Available: https://doi.org/10.4018/978-1-7998-9434-6.ch010

[11] K. Cho et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," arXiv.org, 2014. https://arxiv.org/abs/1406.1078

[12] G. Saggio, F. Riillo, L. Sbernini, and L. R. Quitadamo, "Resistive flex sensors: a survey," Smart Materials and Structures, vol. 25, no. 1, p. 013001, Dec. 2015, doi: https://doi.org/10.1088/0964-1726/25/1/013001.

[13] J. DelPreto, J. Hughes, M. D'Aria, M. de Fazio, and D. Rus, "A Wearable Smart Glove and Its Application of Pose and Gesture Detection to Sign Language Classification," IEEE Robotics and Automation Letters, vol. 7, no. 4, pp. 10589–10596, Oct. 2022, doi: https://doi.org/10.1109/lra.2022.3191232.

[14] L. Fan, Z. Zhang, B. Zhu, D. Zuo, X. Yu, and Y. Wang, "Smart-Data-Glove-Based Gesture Recognition for Amphibious Communication," Micromachines, vol. 14, no. 11, p. 2050, Nov. 2023, doi: https://doi.org/10.3390/mi14112050.

[15] C. Salchow-Hömmen, L. Callies, D. Laidig, M. Valtin, T. Schauer, and T. Seel, "A Tangible Solution for Hand Motion Tracking in Clinical Applications," Sensors (Basel, Switzerland), vol. 19, no. 1, p. 208, Jan. 2019, doi: https://doi.org/10.3390/s19010208.

[16] M. A. Wa'ie Hazman et al., "IMU sensor-based data glove for finger joint measurement," Indonesian Journal of Electrical Engineering and Computer Science, vol. 20, no. 1, p. 82, Oct. 2020, doi: https://doi.org/10.11591/ijeecs.v20.i1.pp82-88.

[17] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 3, pp. 311–324, May 2007, doi: https://doi.org/10.1109/tsmcc.2007.893280.

[18] Auslan Signbank, "Auslan Signbank," in Auslan Signbank. [Online]. Available: https://auslan.org.au/. Accessed on: Mar. 24, 2024.

[19] Rayane Tchantchane, H. Zhou, Z. Shen, and Gursel Alici, "A Review of Hand Gesture Recognition Systems Based on Noninvasive Wearable Sensors," Advanced intelligent systems, Jul. 2023, doi: https://doi.org/10.1002/aisy.202300207.

[20] R. Khusainov, D. Azzi, I. Achumba, and S. Bersch, "Real-Time Human Ambulation, Activity, and Physiological Monitoring: Taxonomy of Issues, Techniques, Applications, Challenges and Limitations," Sensors, vol. 13, no. 10, pp. 12852–12902, Sep. 2013, doi: https://doi.org/10.3390/s131012852.

[21] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," Scientific Reports, vol. 6, no. 1, Nov. 2016, doi: https://doi.org/10.1038/srep36571.

[22] J. Chapman, A. Dwivedi, and Minas Liarokapis, "A Wearable, Open-Source, Lightweight Forcemyography Armband: On Intuitive, Robust Muscle-Machine Interfaces," Sep. 2021, doi: https://doi.org/10.1109/iros51168.2021.9636345.

[23] Z. Liang et al., "A Wireless, High-Quality, Soft and Portable Wrist-Worn System for sEMG Signal Detection," Micromachines, vol. 14, no. 5, pp. 1085–1085, May 2023, doi: https://doi.org/10.3390/mi14051085.

[24] X. Yang, X. Sun, D. Zhou, Y. Li, and H. Liu, "Towards Wearable A-Mode Ultrasound Sensing for Real-Time Finger Motion Recognition," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 6, pp. 1199–1208, Jun. 2018, doi: https://doi.org/10.1109/tnsre.2018.2829913.

[25] S. Khaksar, "A framework for gamification of remote wrist rehabilitation exercises utilising non-invasive sensors," espace.curtin.edu.au, 2023. https://espace.curtin.edu.au/handle/20.500.11937/91409?show=full (accessed Apr. 28, 2024).

[26] Xu Zhang, Xiang Chen, Yun Li, V. Lantz, Kongqiao Wang, and Jihai Yang, "A Framework for Hand Gesture Recognition Based on Accelerometer and EMG Sensors," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 41, no. 6, pp. 1064–1076, Nov. 2011, doi: https://doi.org/10.1109/tsmca.2011.2116004.

[27] Y. Gu et al., "A Review of Hand Function Rehabilitation Systems Based on Hand Motion Recognition Devices and Artificial Intelligence," Brain Sciences, vol. 12, no. 8, p. 1079, Aug. 2022, doi: https://doi.org/10.3390/brainsci12081079.

[28] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A Review on Wearable Photoplethysmography Sensors and Their Potential Future Applications in Health Care," International Journal of Biosensors & Bioelectronics, vol. 4, no. 4, 2018, doi: https://doi.org/10.15406/ijbsbe.2018.04.00125.

[29] T. Zhao, J. Liu, Y. Wang, H. Liu, and Y. Chen, "Towards Low-Cost Sign Language Gesture Recognition Leveraging Wearables," IEEE Transactions on Mobile Computing, vol. 20, no. 4, pp. 1685–1701, Apr. 2021, doi: https://doi.org/10.1109/tmc.2019.2962760.

[30] D. Ferigo, L.-K. Merhi, B. Pousett, Z. G. Xiao, and C. Menon, "A Case Study of a Force-myography Controlled Bionic Hand Mitigating Limb Position Effect," Journal of Bionic Engineering, vol. 14, no. 4, pp. 692–705, Dec. 2017, doi: https://doi.org/10.1016/s1672-6529(16)60435-3.

[31] Y. Yu, X. Chen, S. Cao, X. Zhang, and X. Chen, "Exploration of Chinese Sign Language Recognition Using Wearable Sensors Based on Deep Belief Net," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 5, pp. 1310–1320, May 2020, doi: https://doi.org/10.1109/jbhi.2019.2941535.

[32] J. Brownlee, "How to Choose a Feature Selection Method For Machine Learning," Machine Learning Mastery, Nov. 26, 2019. https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/#:~:text=Filter%20feature%20selection%20methods%20use

[33] B. G. Lee and S. M. Lee, "Smart Wearable Hand Device for Sign Language Interpretation System With Sensors Fusion," IEEE Sensors Journal, vol. 18, no. 3, pp. 1224–1232, Feb. 2018, doi: https://doi.org/10.1109/jsen.2017.2779466.

[34] H. Bello, S. Suh, D. GeißLer, L. Shakti, B. Zhou, and P. Lukowicz, "CaptAinGlove: Capacitive and Inertial Fusion-Based Glove for Real-Time on Edge Hand Gesture Recognition for Drone Control," Oct. 2023, doi: https://doi.org/10.1145/3594739.3610713.

[35] A. Mishra, J. Kim, J. Cha, D. Kim, and S. Kim, "Authorized Traffic Controller Hand Gesture Recognition for Situation-Aware Autonomous Driving," Sensors, vol. 21, no. 23, pp. 7914–7914, Nov. 2021, doi: https://doi.org/10.3390/s21237914.

[36] K. Guo, M. Orban, J. Lu, M. S. Al-Quraishi, H. Yang, and M. Elsamanty, "Empowering Hand Rehabilitation with AI-Powered Gesture Recognition: A Study of an sEMG-Based System," Bioengineering, vol. 10, no. 5, p. 557, May 2023, doi: https://doi.org/10.3390/bioengineering10050557.

[37] R. E. Nogales and M. E. Benalcázar, "Hand Gesture Recognition Using Automatic Feature Extraction and Deep Learning Algorithms with Memory," Big data and cognitive computing, vol. 7, no. 2, pp. 102–102, May 2023, doi: https://doi.org/10.3390/bdcc7020102.

[38] F. Al Farid et al., "A Structured and Methodological Review on Vision-Based Hand Gesture Recognition System," Journal of Imaging, vol. 8, no. 6, p. 153, May 2022, doi: https://doi.org/10.3390/jimaging8060153.

[39] C. Salchow-Hömmen, L. Callies, D. Laidig, M. Valtin, T. Schauer, and T. Seel, "A Tangible Solution for Hand Motion Tracking in Clinical Applications," Sensors (Basel, Switzerland), vol. 19, no. 1, p. 208, Jan. 2019, doi: https://doi.org/10.3390/s19010208. '

[40] Silvera-Tawil, David & Rye, David & Soleimani, Manuchehr & Velonaki, Mari. (2015). Electrical Impedance Tomography for Artificial Sensitive Robotic Skin: A Review. Sensors Journal, IEEE. 15. 2001-2016. 10.1109/JSEN.2014.2375346.