Department of Biomedical Engineering
Silesian University of Technology

PhD Thesis

A data glove with a reduced number of sensors for the recognition of Polish Sign Language letters

Author:
mgr inż. Jakub Piskozub

Supervisor:
prof. dr hab. inż. Paweł Strumiłło

Łódź, 2023

Table of contents

Streszczenie

Budowane obecnie rękawice sensorowe do rozpoznawaniu ruchów rąk podczas komunikacji językami migowymi mają niedostateczną ergonomię, niezadowalającą skuteczność, co wynika często z błędnego zrozumienia istoty języka migowego przez badaczy budujących takie rękawice. Celem niniejszej pracy jest budowa i badania autorskiej rękawicy sensorowej o jak najmniejszej liczbie wymaganych czujników. Badania opisane w niniejszej pracy skupiają się na eksperymentach, z użyciem rękawicy własnego projektu, związanych z zastosowaniem technologii uczenia maszynowego. Dane treningowe zostały pozyskane od 15 uczestników wykonujących zarówno statyczne gesty ręki Alfabetu Polskiego Języka Migowego (APJM) jak i gesty dynamiczne. Eksperymenty z gestami statycznymi polegały na przeprowadzeniu ich klasyfikacji za pomocą 6 klasyfikatorów. Najlepszą skuteczność rozpoznawania gestów na poziomie 99% uzyskano dla klasyfikatora *k*-Nearest Neighbours in Random Subspaces. Za pomocą algorytmu drzewa decyzyjnego z zastosowaniem współczynnika Gini'ego wyznaczono hierarchię wpływu poszczególnych czujników piezoelektrycznych rękawicy na skuteczność klasyfikacji a następnie potwierdzono ją w ponownie przeprowadzonych klasyfikacjach gestów. Dla danych z trzech czujników osiągnięto skuteczność 94%. W kolejnym etapie badań rozpoznawano gesty dynamiczne liter Alfabetu Polskiego Języka Migowego. Z wykorzystaniem sieci neuronowej zbudowanej z warstw splotowych oraz jednostek rekurencyjnych GRU przeprowadzono klasyfikację gestów dynamicznych ze skutecznością 99% dla danych zawierających odczyty ze wszystkich 10 czujników piezorezystancyjnych oraz 6 osi czujnika inercyjnego. Najważniejszy wątek pracy skupiał się na selekcji najistotniejszych czujników. Ze względu na silną zależność korelacyjną sygnałów rejestrowanych z czujników wyznaczenie hierarchii cech metodami statystycznymi, przeprowadzono selekcję metodą brute-force, sprawdzając skuteczność każdej możliwej kombinacji w procesie treningu sieci neuronowej. Otrzymane rezultaty dowodzą możliwości klasyfikacji liter APJM ze skutecznością przekraczającej 98% na podstawie danych jedynie z trzech czujników piezoelektrycznych oraz 6 osi czujnika inercyjnego oraz wskazują na istnienie uniwersalnego dla każdej osoby tercetu takich czujników. Dowodząc tezy o skutecznej klasyfikacji liter APJM z wykorzystaniem mniej niż 5 czujników wykazano, że jest możliwe stosowanie rękawic sensorowych o uproszczonej budowie, co może pozwolić na poprawię ergonomii rękawicy, jej niezawodności i możliwość akceptacji przez osoby z niepełnosprawnościami.

Abstract

Data gloves currently being built for recognizing hand movements during sign language communication have inadequate ergonomics, unsatisfactory efficiency, which is due, among other things, to a misunderstanding of the essence of sign language by researchers building such gloves. The purpose of the present work is the construction and testing of the author's sensor glove with the smallest possible number of required sensors. The research described in this paper focuses on experiments involving the use of machine learning technology. Training data was obtained from 15 participants performing both static hand gestures of the Polish Sign Language Alphabet (PSLA) and dynamic gestures. Experiments with static gestures consisted of classifying them using 6 classifiers. The best gesture recognition performance of 99% was obtained for the *k*-Nearest Neighbors in Random Subspaces classifier. Using the decision tree algorithm with the Gini coefficient, the hierarchy of the influence of the individual piezoelectric sensors of the glove on the classification effectiveness was determined and then confirmed in the re-performed gesture classifiers. For data from three sensors, an accuracy of 94% was achieved. In the next stage of the research, dynamic gestures of the letters of the Polish Sign Language Alphabet were recognized. Using a neural network built from spline layers and GRU recursive units, dynamic gesture classification was carried out with an accuracy of 99% for data containing readings from all 10 piezoresistive sensors and six axes of the inertial sensor. The most important strand of work focused on the selection of the most relevant sensors. Due to the strong correlation relationship of the signals recorded from the sensors determining the hierarchy of features by statistical methods, brute-force selection was carried out, checking the effectiveness of each possible combination in the process of training the neural network. The results obtained prove the possibility of classifying PSLA letters with an accuracy exceeding 98% on the basis of data from only three piezoelectric sensors and 6 axes of an inertial sensor, and indicate the existence of a tercet of such sensors that is universal for each person. Proving the thesis of effective classification of PSLA letters using less than 5 sensors, it was shown that it is possible to use data gloves with a simplified design, which may allow to improve the ergonomics of the glove, its reliability and acceptability for people with disabilities.

List of abbreviations

SL - Sign Language

PSL - Polish Sign Language

PSLA - Polish Sign Language Alphabet

ASL - American Sign Language

BSL - British Sign Language

CSL - Chinese Sign Language

RF - Random Forest

HMM - Hidden Markov Model

SVM - Support Vector Machines

MC-DCNN - Multi Channel Deep Convolutional Network

ANN - Artificial Neural Network

t-LeNet - Time-series counterpart of LeNet architectures for image classification

IMU - Inertial Measurement Unit

RNN - Recurrent Neural Network

LSTM - Long Short-Term Memory

GRU - Gated Recurrent Units

DTW - Dynamic Time Warping

k-NN - k- Nearest Neighbours

GPS - Gesture Progression Scalar

MEMS - Micro-Electro-Mechanical Systems

BCI - Brain-Computer Interface

HMD - Head Mounted Display

VR - Virtual Reality

AR - Augmented Reality

PCA - Principal Component Analysis

LDA - Linear Discriminant Analysis

MRMR - Minimum Redundancy Maximum Relevance

PI - Permutational Importance

XCNN - eXplainable Convolutional Neural Networks

1. Introduction

1.1. Purpose of the dissertation

The purpose of this dissertation is to build data glove equipped with set of sensors intended to study the possibility of its use of rehabilitation and diagnostics applications and recognition of the letters of the Polish Sign Language Alphabet - PSLA - . With its help, it was also decided to investigate the feasibility of reducing the number of sensors in a data glove used to recognize the letters of the PSLA. The work aims to analyze existing solutions, develop a new, more efficient and ergonomic model of the data glove, and apply advanced machine and deep learning techniques to improve the quality of gesture recognition.

1.2. Research motivation and thesis

The obvious choice among assistive devices for people with disabilities is primarily a device that can directly combat the disability itself by acting on the damaged or malfunctioning organ. In the case of deaf and dumb people these are usually hearing aids, cochlear implants or vocoders. However, there are cases when the use of such a system is impossible, insufficient or simply not desired by the person. With such eventualities in mind, a number of technologies have been developed to improve the quality of life and safety of their users. These include devices that assist the visually impaired and actively survey the environment and convert its sounds into light signals or text [1]. Technology of the last two decades has also significantly contributed to improving the quality of life of deaf and dumb people [2]. With the use of smartphones, such people can sufficiently communicate with others who do not know sign languages, and with the spread of video calls, connect at usually any time with a sign language interpreter [3]. However, it is not at all improbable that there is no phone at hand, no Internet connection necessary to make a call to an interpreter on an important matter, or that the sign language speaker does not know any other language. In such a situation, a system based on a data glove that can translate sign language into speech, either in combination with a vision system or in a limited way on its own, can come to the aid. A device that has its genesis in cybernetics and has existed in the wider consciousness since it appeared as a game controller.

Over the past decade, the substantial advancements in Virtual Reality - VR - technology have led to the integration of VR systems into the consumer market [4]. These systems typically consist of a Head Mounted Display - HMD - , often referred to as VR goggles, controllers, and a pair of transceivers. The HMD is a screen positioned at a specific distance from the eyes, creating an immersive experience that feels like an alternate reality, commonly known as virtual reality. This powerful impression has revolutionized the way we perceive and interact with digital environments [5].

For a device that simulates reality so convincingly, it is essential to incorporate interactivity to further augment the sense of real experiences. Presently, the most widespread controllers are joystick-like devices called wands, which combine the natural interaction of human hands with conventional gaming input devices [6]. These wands have become an accepted and instinctive means of control, particularly among younger users who have grown up using similar devices.

Over the years, there has been significant progress in developing intuitive interaction technologies for computers, particularly through computer games [7]. These games often necessitate players to execute multiple actions simultaneously, such as moving in the game space, controlling the camera, and performing specific actions. Moreover, online games add the challenge of performing these tasks under stress, arising from competing players and time constraints.

To understand the next stage in human-computer interaction, one can trace a simplified path of how modern devices are controlled. The user's initial intention, whether it involves movement in a computer game or typing on a keyboard, must first be translated by the brain into a series of hand movements, which are then executed. As such, any interaction with a computer can be represented as a brain-hand-controller-computer chain [8].

An alternative solution to computer control is the Brain-Computer Interface - BCI - [9]. However, this technology is still in its infancy, permitting only the most basic actions, and remains expensive and non-ergonomic. Consequently, it is not expected to become a commercial solution for several years. It is reasonable to assume that the next significant milestone in control technology will involve eliminating the need for controller buttons and focusing on controllers that mimic the natural movements of our hands. This approach is most beneficial when applied to objects with a range of motion akin to humans, such as anthropomorphic manipulators or human hands in VR.

In today's world, where telecontrol is becoming increasingly crucial and robotics is experiencing significant breakthroughs, it is essential to develop devices that enable intuitive control of robotic manipulators while ensuring the required precision of their operation. The human hand appears to be the most suitable tool for this purpose. To make this feasible, it is necessary to replicate the motor-nervous system of the hand, allowing operators to freely execute tasks and experience their effects.

Examining the human hand as a mechanical structure, it has a total of 27 degrees of freedom [10]. Taking into consideration the individual human motor coupling, the human hand becomes an extremely intricate mechanical manipulator. Besides the sophisticated muscular and skeletal systems, the hand is the most densely innervated part of the human body, with significant cortical representation. The human brain allocates substantial resources to interpreting the sense of touch signals. This task is so critical and complex that it necessitates the synchronized utilization of both brain hemispheres due to the requirement for feedback in all hand movements [11].

Devices that track hand movements, like the data glove developed by Sandin et al. [12], make this possible. The data glove enables direct measurement of finger movements and is designed ergonomically, without interfering with the user's natural movements. This feature is crucial when testing various gestures and sign languages. As we continue to explore the potential of human-computer interaction, devices like the data glove will play a pivotal role in bridging the gap between the physical and virtual worlds, paving the way for more sophisticated and intuitive control systems. Since the creation of the first data glove, it is only a matter of time that this technology will be used to study the most complex set of structured hand movements: Sign Language.

Sign language is a key communication tool for deaf and hard-of-hearing people. Data gloves have the potential to facilitate communication between these individuals and the environment, but their size and number of sensors may be a barrier to widespread deployment. The motivation for taking up this topic is the desire to contribute to improving the quality of life of people who use sign language, as well as to accelerate the development of technologies to support such communication.

The potential applications of the data glove as a part of sign language translator are vast and far-reaching, opening up new opportunities for communication and social inclusion. As a real-time sign language translator, the data glove can be instrumental in breaking down communication barriers between the deaf and hearing communities. By converting hand gestures into spoken or written language, it can enable seamless interaction between individuals who rely on sign language and those who are not familiar with it. One significant area where such translator can make a difference is in educational institutions [13]. By providing real-time translation of sign language, it can help deaf students more effectively participate in classroom discussions, group projects, and other collaborative activities. This increased level of engagement can lead to improved academic performance and a more enriching educational experience for these students. In the workplace, these devices can play a vital role in fostering an inclusive environment for deaf and hard-of-hearing employees. By facilitating communication with their colleagues and supervisors, it can help them become more integrated into the workforce and contribute more effectively to their teams. A translator can also provide invaluable support during meetings and presentations, ensuring that all employees have equal access to the information being shared. This can be particularly important in remote work settings, where effective communication is essential for team collaboration and project success. Outside of formal settings, the data glove can help promote social inclusion for the deaf community by enabling them to communicate more easily in social gatherings and public spaces. With the ability to translate sign language into spoken or written text, deaf individuals can participate in conversations and engage with others without the need for a human interpreter. This increased level of interaction can lead to a greater sense of belonging and reduced feelings of isolation. One of the most exciting possibilities for the data glove is its potential integration with smartphones or other portable devices. This could enable on-the-go translation, empowering deaf individuals to communicate more easily while traveling or navigating public spaces. Additionally, the data glove could be used in conjunction with Virtual Reality or Augmented Reality - AR - technologies, allowing users to interact with virtual environments using sign language. The data glove can also function as a sign language learning tool, providing real-time feedback on the accuracy of the user's hand gestures and assisting in the development of proper sign language skills. By incorporating haptic feedback, the data glove can help users learn the correct force and pressure required for specific gestures, improving the overall quality of their sign language communication. In the realm of media accessibility, the data glove can significantly enhance the experience for deaf and hard-of-hearing viewers by providing real-time sign language interpretation for live events, television programs, and films. This increased access to media content can lead to a more diverse and inclusive entertainment landscape. Another intriguing application of the data glove is its potential to facilitate communication between speakers of different sign languages. By translating gestures from one sign language to another, it can foster cross-cultural understanding and promote more meaningful interactions between individuals from diverse linguistic backgrounds.

But the most significant of the many uses of the data glove as a sign language interpreter may be its use in emergency situations. The data glove can prove invaluable to first responders, enabling them to communicate more effectively with deaf or hard-of-hearing people who require assistance. By providing real-time sign language translation, the glove can help ensure that key information is accurately communicated and understood, ultimately leading to a more effective emergency response and potentially saving lives. An obstacle to their widespread use, however, is their high cost and unreliability, linked directly to their size as mentioned by their potential users. Reducing the number of sensors used for effective gesture recognition in sign language communication will contribute to solving both of these problems.

The thesis of the following work can hereby be formulated as follows:

> **It is possible to effectively recognize the letters of the Polish Sign Language Alphabet using a sensor glove with a reduced number of sensors, which can simplify the construction of the glove and offer its improved ergonomics.**

1.1. Structure of the thesis

The dissertation consists of six chapters and appendices. Chapter two presents a literature review of the basics of Polish Sign Language, existing data glove solutions, and machine and deep learning technologies used in gesture recognition. Chapter three describes the research methodology, including study group selection, measurement methods, data analysis and feature selection, and the application of machine and deep learning in PSL letter recognition.

Chapter four presents details of the design of the data glove, including the selection of materials and technology, optimization of the glove design, integration of the sensor system, and implementation of gesture recognition algorithms. Chapter five is devoted to evaluating the performance of the developed data glove, including experiments with a control group, assessing the quality of letter recognition, comparing it with existing solutions, and analyzing potential limitations and challenges.

The sixth chapter contains conclusions, including a summary of the research results, contributions, practical implications and prospects for further research. The paper concludes with a bibliography and appendices that include a description of the experiments, implementation source code and detailed experimental results.

2. Theoretical and technological background

2.1. Polish Sign Language: syntax and alphabet

Sign language is a visual-spatial communication system used by deaf and hard of hearing people. The syntax of sign language varies from region to region, as each country has its own independent sign language - e.g. Polish Sign Language, American Sign Language - ASL - , British Sign Language - BSL - , etc. - . The syntax of sign language can be outlined in general terms:

1. Hand configuration: The shape, position of the fingers and position of the hands are key elements of sign making. Different hand configurations represent different words or concepts.
2. Location: The location of the hand in the space in front of the speaker's body affects the meaning of sign. For example, the location of a sign on the face may signal a question, while the location on the arm may signify an adjective.
3. Movement: the movement of the hands, arms and body also affects the meaning of snapshots. Movement can change in speed, direction and distance, which can affect the perception of a sign. The speed and rhythm of gestures can affect intonation and emphasis in sign communication. Slowing down the pace can suggest the importance of the message, while speeding up can indicate excitement or haste.
4. Facial expression: Facial expression is an important part of sign language because it conveys emotions, intonations and additional information related to the snapshots. Smiling, furrowing one's brow, raising one's eyebrows - all can affect the interpretation of sign. Maintaining eye contact with an interlocutor is important in sign communication because it facilitates interpretation of gestures and also

expresses attention and interest. Avoiding eye contact can indicate uncertainty, misunderstanding or unwillingness to talk.

5. Body Position: Posture and body position can affect how sign communication is perceived. A confident and upright posture may indicate confidence and competence, while a hunched posture may suggest uncertainty or lack of commitment. Movements of the arms, torso and legs can reinforce the message of sign language, such as by emphasizing emphasis, energy or intensity of feelings. Gestures can also help convey spatial or directional information.

It is worth remembering that specific rules may vary depending on the sign language in question.

One part of PSL is the sign alphabet, also known as finger sign. It is mainly used to articulate proper names, geographical names, technical names, first and last names and other names that do not have their own sign in sign language

In practice, the finger alphabet facilitates communication between deaf people and between deaf and hearing people who are not fluent in sign language

The finger alphabet is also used as an educational tool for sign language learners. It helps them memorize gestures and letters, as well as achieve fluency in communication. It is worth noting that different countries have different versions of the finger alphabet that correspond to the letters of the language.

Before moving on to a discussion of the PSL Alphabet , it is necessary to explain the difference between a sign and a gesture, and what is meant by dynamics in both cases.

In the context of sign language grammar and syntax, a sign means some specific, structured hand movement that has a meaning defined by the sign language dictionary. It refers to a specific word, concept, phrase or grammatical element that is represented by a specific hand gesture and, in some cases, facial expressions, body position or arm movements. It can be static or dynamic. Dynamic, in this case, means that the essence of the whole sign is represented by a specific movement, such as making a motion with the hand in an arc.

A graphic illustration of the difference between a static and dynamic gesture can be seen in Figure 2.1.1

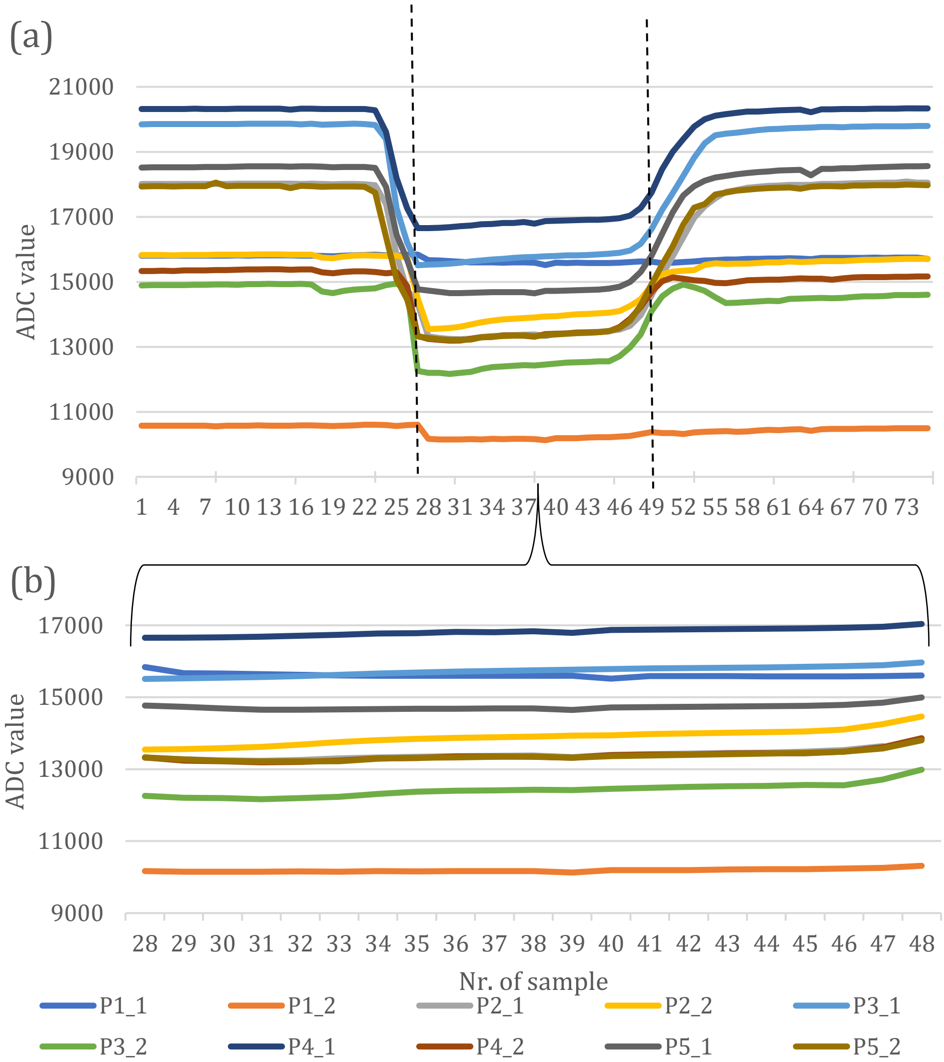

Figure 2.1.1. Visualization of the difference between static and dynamic gesture with visualization of the difference between dynamic - a - and static gesture - b - .

The graph above highlights the readings from the glove sensors during the making of the letter 'A' of the PSL alphabet. Chart - a - shows all three stages of the sign: construction, merit and deconstruction. Construction and deconstruction are the stages in which the fingers from a neutral position fold and unfold into a given mark, respectively. PSLA's letter 'A' is a static letter and its merit is fingers clenched into a fist. Analyzing the sensor signals only at this point, we are dealing with a static gesture. The signals seen in graph - b - are close to constant in time. A static gesture has no preceding dynamic intermediate stages. This is precisely the difference between a gesture and a static sign. Thus, it should be noted that all PSLA letters, even the static ones, are dynamic gestures.

This difference is important in that many researchers conducting work on systems potentially acting as sign language interpreters test their systems precisely on static gestures erroneously calling them precisely static sign alphabet letters. However, it is possible to analyze the individual phases of such letters, as has been done in the following work. In that case, however, such gestures should not be called PSLA letters. Such situation is analogous to trying to call the letter 'V' the letter 'M'.

The PSL finger alphabet consists of 36 characters: 20 dynamic and 16 static. An interesting feature of the APSL is that diacritical marks such as 'Ł' and 'Ó' are the dynamic equivalent of the Latin alphabet characters. All 36 APSL letters are shown in Figure 2.1.2.

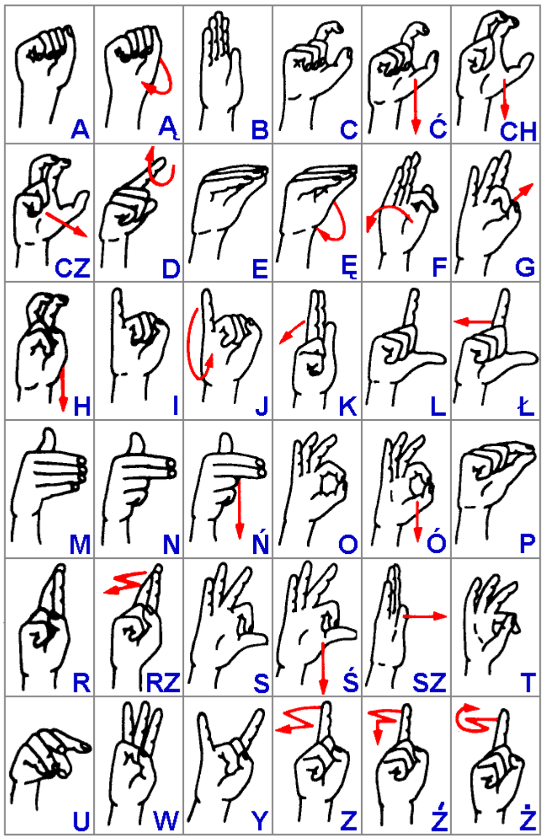

Figure 1.1.2. Polish Sign Language Alphabet source:http://alfabet-migowy.blogspot.com/2013/04/alfabet-migowy.html

And while hand movements make up the majority in SL grammar, they do not make up its entirety. Equally important as hand positioning during communication are facial expressions and body language. So without an additional device that tracks and recognizes in tandem with the glove action of these other two aspects of SL, it is impossible to create a complete interpreter. Thus, it should be made clear that the glove itself will never be a stand-alone sign language interpreter.

2.2. The standpoint of sign language users

It is this misconception about the glove as a complete sign language interpreter that makes, in the world of technology, many sign language users express negative opinions about various devices and applications that serve as interpreters. Although many of them are being developed to facilitate communication between deaf, hard-of-hearing and hearing people, there are still significant problems that make these technologies far from perfect.

First of all, most of the researchers developing sign language interpreters are hearing people. They often mistakenly assume that a sign language interpreter in the form of data gloves will be a complete device, but in reality it does not reflect all aspects of sign language. As mentioned earlier, sign language uses not only hand movements, but also facial expressions, localization and other elements of non-verbal communication. Data gloves focus only on hand movements, leading to incomplete and inaccurate translation.

Another problem is the lack of consideration of context in translation. Sign language is based on context and situation, and interpreters are not always able to distinguish nuances of meaning. As a result, people who use these devices can feel frustrated by misunderstandings and difficulties in communication [14].

In addition, many sign language interpreter applications and devices do not take into account the regional differences and peculiarities of individual sign languages. Therefore, a person using Polish Sign Language may experience communication problems when the app is adapted for ASL [15].

Some sign language users also believe that these technologies may detract from the value of sign language learning for hearing people. Instead of encouraging education and understanding of deaf culture, these devices may make hearing people less motivated to learn sign language, which may lead to further alienation of deaf people.

Many of the devices created as components of such a translating system to track hand movements are of an unergonomic size. It is not uncommon for these devices to be devices whose components extend to the user's forearm, connecting to the sensors on the hand via a tangle of wires. It is not surprising, then, that the sight of such an unergonomic device arouses resentment among those who would potentially carry it with them every day. The large size would not only make the device difficult to use, but would also expose it to damage and thus misinterpret the SL and thus exacerbate the language barrier.

While there is a lot of negativity about sign language interpreters, it is also worth noting that developments in technology may lead to improvements in these tools in the future. As machine learning and artificial intelligence develop, it is possible that interpreters designed for sign languages will become more accurate and useful. By leveraging machine learning algorithms, the data glove can continually improve its translation accuracy over time, adapting to the nuances of various sign languages and individual users' signing styles. This adaptability can lead to a more accurate and personalized translation experience, ultimately enhancing the overall effectiveness of a sign language translator.

2.3. Data gloves: methods and existing solutions

It is first necessary to identify the fundamental problems associated with the issue of recognizing gestures and hand movements in sign languages.

The challenge of hand gesture recognition can be broken down into three primary components - Fig. 2.3.1 - : the measurement unit comprised of various sensor types, the measurement data processing block that conditions and formats the data for presentation to a data classifier, and finally, the data classifier, which is initially trained on the training data and subsequently used for classifying new data into one of the predefined classes representing individual signs.

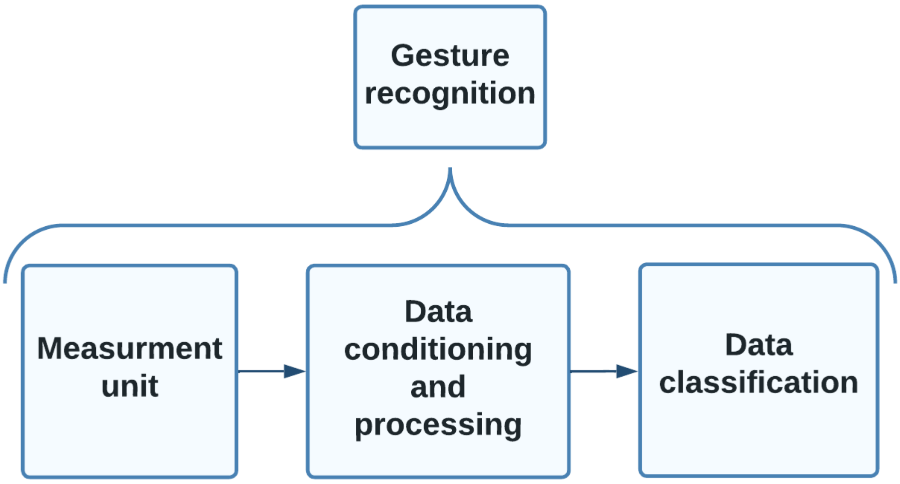

Figure 2.3.1. Division of problems of hand gesture recognition issue

In general, a division of methods for tracking finger and whole hand movements can be adopted due to the location of the measuring device relative to the hand. Either it is located away from the hand or directly on it:

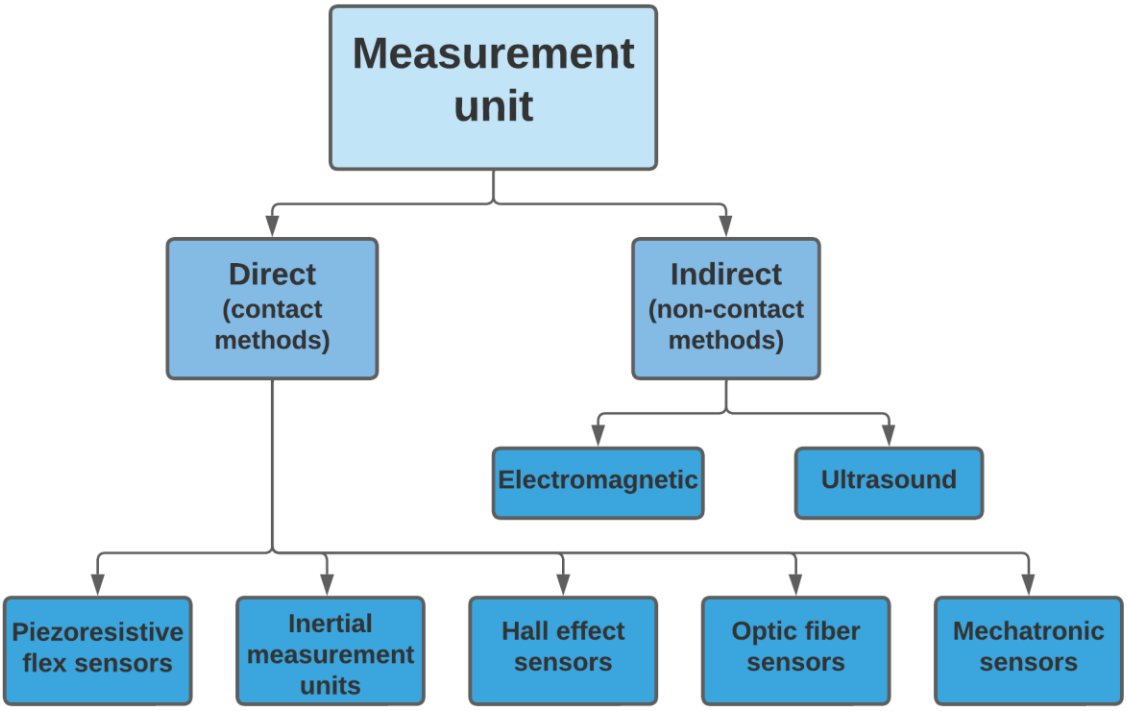

Figure 2.3.2. A breakdown of the types of hand tracking methods

Indirect measurement techniques involve the use of electromagnetic waves and sound waves. The most prevalent indirect method employs cameras to track hand movements through image analysis. According to the Authors of [16] who summarized more than 100 articles on sign language interpretation and recognition, 81% of the systems rely on vision sensors. In addition to cameras, which make up 55% of the basis of all the articles collected by the authors, Kinect - 20% - and Leap Motion - 6% - devices are popular. These are devices that have sensors for electromagnetic waves of different spectrums. Tracking of hand movements in their case is based on the synthesis of data received from the sensors.

In [17] Rastgoo et al. review multiple vision methods to solve this problem. They present a breakdown of the methods by type of vision: RGB and depth. The authors summarize dozens of works describing different computer vision models for sign language recognition. According to this summary, depth cameras are the most commonly used. Among the collected works, static images are the most commonly used for recognition. Many of the presented methods achieve 90% accuracy, reaching up to 99.8%. The vast majority of the tests are carried out under laboratory conditions that provide adequate lighting and framing. Nevertheless, many of the presented methods, despite their high efficiency, are not resistant to occlusion.

Nevertheless, image-based hand tracking is subject to constraints due to the inherent nature of the measurement method. The camera is unable to capture movements that occur outside its field of view or those obscured by another object. In contrast, the other types of sensors require the hand to be within their specified field of operation.

Numerous everyday gestures or hand movements, such as shaking hands, may result in the skeleton tracking algorithm being incapable of determining the hand's position. Those issues does not arise with hand tracking using data gloves [18], [19].

Another significant disadvantage of systems based on computer vision is the high dependence on lighting conditions and the frame. The object under examination must be in the right position relative to the camera and in the right lighting. This is difficult to ensure in real-world conditions, especially in emergency situations.

Glove-based devices - direct methods of tracking - that, according to the [16] account for 8% of works on sign language recognition, are free from the problem of occlusion, which, along with independence from lighting conditions are the biggest advantages of that approach.

Direct methods are divided according to the type of sensors used. According to [20] approx. 85% of the solutions collected in the study are based on piezoresistive sensors. The sensors take the form of flexible strips on the surface of which a conductive suspension containing carbon particles has been sprayed. As a result of bending the sensor, the distance between the individual particles expands, thereby increasing the resistance of the sensor.

Optical sensors, already used in Robert Sayre's glove [12], are based on a similar principle as piezoresistive ones. They consist of an optical fiber placed over the finger joints, and a light emitter and detector placed at opposite ends of the fiber. As a result of the bending of the finger, the optical fiber placed over it is also bent, which consequently leads to the escape of a part of the beam's energy of a magnitude proportional to the bending angle of the finger, due to the change in the angle of incidence of the beam inside the optical fiber structure. The difference between the emitted and received beam is measured using a photodiode at the other end of the optical fiber.

The second popular type of sensor is Micro-Electro-Mechanical Systems - MEMS - inertial sensors. MEMS inertial sensors are miniature devices that measure linear acceleration and angular velocity. They use microscopic mechanical elements, such as reference masses and vibrating structures, which are displaced by forces applied to them as a result of motion. These displacements are measured using various measurement techniques and then converted into electrical signals that are proportional to the values of acceleration and angular velocity. MEMS inertial sensors are used in many applications such as navigation, stabilization, motion measurement and motion control. In gloves, they are most often placed on the tips of the fingers and in the center of the glove. By comparing the accelerations and angular velocities of individual sensors, it is possible to determine their position relative to a reference one placed on the middle part of the hand and thus the flexion of the fingers.

Finger tracking methods using Hall effect sensors operate on a similar principle. They work on the basis of a phenomenon discovered by Edwin Hall, which involves generating a voltage - Hall voltage - transverse to the flow of current in a conductor or semiconductor when the material is subjected to a magnetic field perpendicular to the direction of current flow. These sensors use the Hall voltage to measure the magnitude of magnetic fields or electric currents. When a Hall effect sensor is placed in a magnetic field, a Hall voltage proportional to the intensity of the magnetic field is generated. This voltage is measured and converted into an electrical signal that can be used for various applications, such as measuring speed, position, rotation or current flow. Compared to solutions based on inertial sensors, the biggest disadvantage of such devices is the need to place the hand in a magnetic field of a certain area [21].

There are also mechatronic methods. The most common are systems of tendons attached to finger tips on one side and a rotating element on the other: whether to the rotor of a DC motor working as a generator in this case, or to the shaft of an encoder or potentiometer measuring the rotation of the tendon spool. Such solutions, however, are found in a small number of cases. They were used in prototypes of orthopedic devices in the last century. Currently, such a system can be used by the LucidVR project, a low-cost haptic glove for VR systems that is an open project [22].

The advantage of piezoresistive sensors over any of the other indirect methods of measuring finger placement is primarily the simplicity of application. In a measuring circuit, they can act as a potentiometer with a bias proportional to the bending of the finger. More technical characteristics of piezoresistive sensors will be described in Chapter 4.

An additional advantage of such sensors is the relatively low price relative to other solutions. Moreover, in [23] researchers have developed a simple and inexpensive stretchable piezoresistive sensor based on a blend of black carbon and Ecoflex that can be mass-produced. The sensor has a unique zig-zag pattern that allows it to measure the flexion angles of the proximal interphalangeal and metacarpophalangeal joints. The sensor has proven sufficiently effective for use in a data glove. However, it is worth noting some limitations related to the slightly different behavior of elastomer-based sensors, such as stress relaxation and changes due to different strain rates. Dopracowanie tej metody może pozwolić na dalszą redukcję kosztu czujników piezorezystancyjnych do rękawic danych.

In [24], [25] Lee et al. propose a different approach to implementing piezoresistive sensors. The solution described by the authors involves making sensors via direct writing of eutectic Gallium Indium - eGaIn - on the surface of a glove.

2.4. Machine and deep learning technologies applied to gesture recognition

Numerous articles published in recent years regarding data gloves have emphasized their abilities in identifying Sign Language - SL - signs or gestures. It is crucial to highlight the distinctions between SL signs. The sign gestures can be categorized based on the dynamics of a particular gesture. Static signs are simpler to detect since even when presented in a sequence, there is a natural pause point in each sign, i.e., when the hand is motionless. Conversely, recognizing dynamic signs is more difficult due to the absence of such a point. Furthermore, similar to speech, the spacing between each sign in SL is vital for the identification of individual signs. An additional aspect of recognizing and classifying sign alphabet letters is that they should always be interpreted as dynamic gestures, and therefore as time series. After applying appropriate feature engineering, it is possible to classify time series using machine learning algorithms, but there are also algorithms dedicated to time series classification. A few of the most commonly used are described in the following.

The k-Nearest Neighbors - k-NN - algorithm is a supervised learning method used in classification and regression problems. It is based on the assumption that similar objects occur close to each other in the feature space. For a given object, the k-NN algorithm assigns a label or value based on information about nearest neighbors from a learning set. Data preparation is the first step in the k-NN learning process. A learning set with labels - for classification - or target values - for regression - must be prepared. Then, an appropriate distance metric is chosen to assess the similarity between objects. Popular metrics include Euclidean, Manhattan or Minkowski distance. Feature normalization is also an important step, especially when individual features have different ranges of values. Normalization avoids the dominance of one feature over others when calculating distances. The next step is to choose the value of k, the number of nearest neighbors to be included in the algorithm. The optimal value of k depends on the data and the goals of the analysis. Once all these parameters are set, you can proceed to determine the nearest neighbors for each sample from the test set. It is necessary to calculate the distances between this sample and all samples of the learning set, and then select the k samples with the smallest distances as nearest neighbors [26].

K-NN in random subspaces - k-NNiRS - involves using the k-NN algorithm in combination with the random feature subspace technique. This technique is often used as a way to increase model accuracy and stability and reduce the impact of overfitting. In k-NN in random subspaces, the learning and classification process is solved by using multiple k-NN classifiers, each working on a subset of the input features. Each of these classifiers is learned on a different subset of the features of the original data set During the learning process, for each k-NN classifier, a subset of features is randomly selected from the full feature set. Then, the k-NN classifier is learned on this subset of features. Feature selection is performed independently for each k-NN classifier. During the classification process, for each sample from the test set, each k-NN classifier makes a prediction based on the previously selected features. Eventually, these predictions are combined to arrive at a final decision, for example through majority voting. The use of k-NNiRS can lead to better results, as reducing the dimensionality of the data reduces the impact of noisy, irrelevant or correlated features. In addition, the diversity of classifiers can increase the stability of the model, making it more resistant to different types of disturbances. In practice, this means that the method can lead to better results compared to a single K-NN classifier [27].

The Support Vector Machines - SVM - algorithm is a machine learning method mainly used in binary classification tasks, although applications in regression and multi-class classification are also possible. SVM is based on the idea of finding the hyperplane that best separates two classes of data in the input space. The most important aspect of SVM is the concept of margin - the distance between the hyperplane and the closest data points of the two classes. The SVM seeks to find the hyperplane that maximizes the margin - the distance between the hyperplane and the closest points from both classes. These points are called support vectors and are crucial to the learning process. The input data is transformed to the input space using kernel functions. These functions allow working with data that is not linearly separable in the original feature space. With kernel functions, it is possible to find a separating hyperplane in a higher-dimensional space. Popular kernel functions include linear, polynomial, radial or sigmoidal. In the learning process, the SVM algorithm solves the optimization problem of maximizing the margin while minimizing the penalty for misclassifying points. The value of the penalty is controlled by the parameter C, which governs the balance between margin and misclassification errors. For multi-class classification, the SVM algorithm is typically used in a one-vs-one or one-vs-all scheme. For regression, the SVM is modified to minimize regression error with an added penalty for exceeding a fixed error threshold. In practice, the SVM algorithm is very effective in classification tasks, especially when the data are linearly separable or require transformation to a higher dimension. In addition, SVM is resistant to over-fitting and generalizes well, even for small learning sets. However, its computational complexity can be a problem for very large data sets [28].

The Hidden Markov Model - HMM - is a unique approach to problems where sequential or temporal dependencies are key. Unlike previously mentioned methods that tend to treat the data as independent and identically distributed, the HMM takes into account the temporal structure of the data. HMMs are statistical models that operate on 'hidden' states - states that are not directly observed, but produce a sequence of observable outcomes. Each transition between states is dependent on the previous state, making it possible to model temporal dependencies. For example, in DNA sequence analysis, the 'hidden' state may be the true function of a DNA segment, while the observable results may be nucleotide sequences. On the other hand, methods such as k-NN and SVM are more traditional machine learning methods. The k-NN algorithm is based on the idea that observations are similar to their nearest neighbors in feature space. SVM, on the other hand, tries to find a hyperplane that optimally divides the data into two or more classes. Both of these methods require a training data set to 'learn' the model, and then apply the model to predict unknown data. While these algorithms are effective for a variety of problems, they can run into difficulties in situations where the temporal relationship between the data is significant - a strong point for HMM [29].

The Dynamic Time Warping - DTW - algorithm is an advanced sequential analysis technique that is extremely useful for comparing sequences of different lengths and rates. Opposing to algorithms such as k-NN, HMMs or SVMs, which generally assume a fixed length and a fixed rate of sequences, DTW allows the comparison of sequences that may be accelerated or decelerated at different locations. DTW accomplishes this by finding the shortest path - least cost - between the beginning and end of two sequences, allowing 'time distortion' in both sequences to achieve the best match. While k-NN and SVM algorithms are effective in classifying data points in feature space based on static data structure, they do not take into account sequence dynamics. HMM, while taking into account the temporal structure of the data, nevertheless assumes a constant rate of sequences and does not deal well with sequences of different lengths. DTW, on the other hand, is flexible in both length and sequence rate. This characteristic makes it particularly useful in applications such as speech recognition or biomedical signal analysis, where sequences can have natural variability in tempo and length [30].

Another alternative is the use of neural networks. The advantage of this machine learning approach is that networks generally allow much faster prediction at the expense of the time needed to train the algorithm. Static gestures can be analyzed using simple network architectures such as the Multi Layered Perceptron. Classification of dynamic gestures is a more complex task because dependencies between samples taken at successive discrete time intervals must be taken into account. Recurrent networks such as LSTM and GRU are suitable for this purpose.

Long Short-Term Memory - LSTM - and Gated Recurrent Units - GRU - are two popular types of Recurrent Neural Networks - RNNs - that are designed to better model sequential data compared to standard RNNs. Both networks, LSTM and GRU, introduce the concept of "gates" into the model to regulate the flow of information through the sequences. LSTM networks, originally proposed by Hochreiter and Schmidhuber in 1997, use three different gates - a forgetting gate, an input gate and an output gate - to control the information flowing through the network. These gates allow the LSTM network to teach long-lived relationships, while reducing the problem of fading gradients that is common in standard RNNs. On the other hand, GRUs, introduced by Cho et al. in 2014, simplify the LSTM structure by reducing the number of gates to two: a reset gate and an update gate. Despite this simplified structure, GRU often achieves results comparable to LSTM in many tasks, while reducing computational complexity and memory requirements. Comparing the two structures, one major difference is that LSTM is better able to model long-term dependencies due to the forgetting gate, which is absent in GRU. On the other hand, GRU tends to be easier to train and more computationally efficient, making it more attractive in applications where computational resources are limited [31], [32].

2.5. Data gloves: related work

The previous section presented recent developments in the field of the sensors themselves used in general-purpose data gloves. However, it is appropriate here to look at the applications of data gloves and the sensors themselves used in them for sign language recognition.

In [33] Simoes Dias et al. presents the development and analysis of a pattern recognition system for the Brazilian Sign Language - Libras - alphabet, based on a glove with sensors. The glove developed consists of five flex sensors, an accelerometer, a gyroscope and two touch sensors. These tactile sensors, made of fabric, have the advantage of not interfering with the opening, closing and movement of fingers.

In the proposed data segmentation method, gestures were divided into three windows that represent the gesture creation period, gesture period and gesture relaxation period. Only the gesture period, which is the target of this analysis, was considered for classification.

With the segmented data, gesture recognition was carried out using different classifiers in different scenarios to select the best features, identify the importance of each type of sensor, select the best combination of sensors, analyze the results for each volunteer and the effect that the proportion of data separated for training has on classification

The sensor analysis succeeded in identifying the most relevant group of sensors for classifying the proposed gesture vocabulary. Based on data from the flex sensors only, a classification accuracy of 86 was obtained using the Random Forest - RF - algorithm. The sensor group consisting of flexion sensors, accelerometer, gyroscope and touch sensors achieved the best results in letter classification - 96.15% accuracy with the RF classifier - .

The proposed system can be used alone or in combination with other types of sign language gesture recognition methods. Future work includes recognizing other Libras' gestures - i.e., words, vocabulary and sentences - , testing with a larger group of volunteers, detecting the beginning and ending moment of each gesture for a real-time automatic detection system.

Lee et al. present the design and implementation of a portable solution for word-based ASL interpretation by analyzing finger and hand movement patterns based on motion data from 5 Inertial Measurement Unit - IMU - sensors [34]. The Recurrent Neural Network model with LSTM layer was tuned to achieve the best performance in classifying 27 word-based ASL gestures, with an average accuracy exceeding 99%.

Although the current study only considered one-handed word-based ASL, preliminary results are promising for further research. Future work includes extending to two-handed word-based ASL and potentially to sentences in ASL - and other sign languages - , as well as applying an automatic algorithm for tuning network hyperparameters to optimize the network.

In [35] Pezzouli et al. presents the process of translating dynamic characters using a specially designed data glove, called Talking Hands. Various classifiers were tested, and some of them, especially Random Forest and neural networks. The study shows that high scores in sign language recognition can be achieved without the use of an external camera or position tracking, acquiring all the data from the wearable device. Indeed, even if specific hand position information is not available, orientation data allows for high translation accuracy. The authors of the article report a gesture classification effectiveness of 99.7% using a neural network based on LSTM cells, based on data acquired from 5 people. The glove constructed by the authors was created using 10 piezoresistive sensors and one inertial sensor.

The results presented here suggest that it is possible to create a fully portable sign language translation system, although several steps are still missing to achieve this goal. First, the segmentation method used to isolate dynamic signs in our study might not be a feasible solution for real-world applications. In fact, users should be asked to input breaks before and after each character to achieve good accuracy in the segmentation process.

The answer to the problem of segmenting a data stream containing gestures was presented in [16] Bae et al. This article presents a real-time sign language recognition system that uses a sensor glove based on 10 piezoresistive sensors sprayed directly onto the glove material. The developed gesture recognition system aimed to segment and recognize 17 dynamic gestures in real time, regardless of the number of repetitions during the gesture. The group of gestures studied included unique gestures, i.e. those that consist of movements that are unique in their length, and repetitive gestures, i.e. those that consist of repetition of shorter and simpler movements. The system consists of three stages: gesture detection, sequence compression and gesture recognition.

At the first stage, the gesture detection model predicts the value of Gesture Progression Scalar - GPS - , which has been proposed for quantum expression of gesture progress. Because the proposed gesture detection method reflects both the dynamics and shapes of hands during gestures, it was possible to achieve accurate gesture detection. In the second stage, all repetitive movements in a given gesture pattern are removed using a repetition removal algorithm, and the lengths of gesture patterns are further shortened using a sequence simplification algorithm. By using the proposed sequence compression stage, the gesture recognition process was accelerated by 80% for repetitive gestures and by 55% for unique gestures. In the final stage, the compressed gesture pattern is classified by the gesture recognition model. In a 13-fold cross-validation test, an accuracy of 94% was achieved by the gesture recognition model.

After separately validating each stage, the three stages were combined into a real-time gesture recognition system and its real-time performance was tested. Since the combined system took only 28 ms from detecting the end of a gesture to updating the recognition result, the user's impression was that the proposed system recognized the gesture immediately.

The research by Saggio et al. [36] describes the use of a glove re-equipped with a set of 10 piezoresistive sensors placed in fabric pockets and one inertial sensor. The resulting data were classified using a nonparametric model combining a k-NN classifier with a DTW algorithm. The model showed good classification accuracy - 96.6% - , with a time complexity that increases linearly with the number of examples used to find k-nearest neighbors for classification. It was observed that the model's classification accuracy improved as the number of training examples increased, but this was at the expense of increased classification time.

Therefore, for practical applications of sign language classification, especially in real time, using the k-NN and DTW model demands a trade-off between accuracy and processing time. Despite the longer classification time being a significant drawback for this model when dealing with a large dataset, its high accuracy is a factor to be considered, especially in scenarios where the number of examples for the classification is limited.

As expected, the computing time of this model, as with other models, highly depends on the hardware used to run it. Given the relatively high classification accuracy and linear time complexity, the k-NN and DTW model could be a good option for applications where processing and storage capabilities are limited and the number of available data is modest. On the other hand, the CNN parametric classifier performed with a very high accuracy, approaching 100%. Thus, the authors note the potential of convolutional neural networks in the problem of time series recognition

Solutions based on devices with fewer than 5 sensors should also be mentioned.

In [37] Gupta et al. describe their study considered the classification of continuously signed sentences using data from one IMU placed on each hand of the signer. The accelerometer and gyroscope signals recorded were processed to generate position data, which were subsequently used as input to three deep-learning models for classification. The set of gestures studied consisted of 15 basic, common Indian Sign Language signs.

Two Convolutional Neural Network - CNN - models for time-series classification, specifically Multi Channel Deep Convolutional Netowork - MC-DCNN - and time-series counterpart of LeNet architectures for image classification - t-LeNet - , were implemented for the classification task. The MC-DCNN, being the deepest architecture, achieved the highest overall accuracy of 83.94%. The t-LeNet model, on the other hand, being a relatively shallow model, resulted in the least classification accuracy of 79.70%. Despite the use of data augmentation with t-LeNet, the model displayed signs of overfitting.

Consequently, a modified t-LeNet architecture was proposed, featuring a reduced number of neurons in the fully-connected layer and an increased number of filters in the convolution layer. This modified t-LeNet architecture not only improved generalization, but also yielded a better classification accuracy of 81.62%, outperforming the original t-LeNet model.

A completely different application of the data glove is presented by Devnath et al. in [38]. In this study, it was found that among the two main algorithms frequently used in the field of posture recognition, the Random Forest-based classifier outperformed the neural network-based classifier in hand posture recognition. The classifiers were evaluated and compared based on several metrics, such as true positive rate, false positive rate, precision, completeness, F-measure and ROC area. Confusion matrices were also prepared based on the data sets provided to the classifiers.

According to the experimental results, classifiers based on neural networks and Random Forest achieved an accuracy of 69% and 97%, respectively. Using these classification results, a basic mapping of human hand postures to robot hands was attempted.

It is possible to improve the accuracy of posture recognition by integrating more sensors into the data glove to cover all fingers. In addition, the use of Hall effect sensors on the fingertips of the robot hand could enable control of grip strength.

A description of the research conducted by Zhang et al. in [39] presents a novel wrist-worn armband based on pressure sensors for real-time hand gesture recognition. Experimental results showed that the armband was able to classify both static and dynamic gestures in real time, and showed good robustness under different scenarios such as re-wearing, different hand positions and different users. The final effectiveness reported by the authors was 90%.

A limitation of this study is the small number of pressure sensors. Although the larger sized sensors used have better robustness for gesture recognition in different scenarios, using a small number of them leads to a loss of wrist pressure information and a reduction in the number of recognizable gestures.

Authors of [40] by sewing Reduced Graphene Oxide - RGO - -coated fiber onto the glove, have designed and produced a kind of flexible data collection glove that has the distinct advantages of simplicity, low cost and comfortable wearing. More importantly, thanks to the good linearity, repeatability and fast response of the RGO-coated fiber to a small tensile stress, they successfully used the manufactured glove to monitor the flexion angle of the finger joint in real time, and then realized the recognition of static and dynamic gestures. The recognition accuracy of ten static gestures representing the digits "0" to "9" is 98.5% in 2000 test samples, which shows the good stability and repeatability of the glove. In addition, experimental results on the recognition of nine words in Chinese Sign Language - CSL - demonstrate good real-time performance where accuracy reaches 98.3% in 180 test samples.

Previous research on PSL letter or gesture recognition from glove data has been limited to small sets of classified hand placements.

In [41] Pławiak et al. described an attempt to analyze 22 hand gestures derived from generally accepted body language for identification. The experiment used data from a specialized glove - with ten sensors - , which was pre-processed and analyzed using machine learning algorithms. The results confirm that efficient and fast recognition of hand body language is possible, with the best classifier achieving a sensitivity of 98.32%.

Dziubich et al., authors of [42], conducted a study on the effectiveness of PSLA letter classification using a glove consisting of 5 flex sensors. In the experiments conducted, which included tests on static and dynamic gesture classification, the ANN classifier performed worse than the SVM. The results for both classifiers can be considered good, with worst-case achievements of 87% and 82%, respectively. The dynamic gesture classifier showed that the performance of the DTW and Hidden Markov Models - HMM - methods was significantly worse, at 48% and 53% respectively. The authors also did not specify which gestures or letters their study was ultimately concerned with. It is only known that there were 25 different static and dynamic gestures.

Currently, to the best of the authors' knowledge, the only research on the PSLA letter recognition glove is conducted by Korzeniewska et al. described in [43]. The article describes the process of developing and testing a prototype sign language translation glove. The prototype allows users to enter letters using a mobile device or computer, replacing a keyboard, although the letter identification process is not fully accurate. The authors also focus on the challenges associated with the durability of thin conductor layers and the longevity of textronic sensors such as Velostat, proposing to use them to make finger flex sensors. An analysis of the electrical properties and selection of substrates for manufacturing thin-film sensors made it possible to create a glove that has been used to classify more than 500 sign samples with a correctness rate of 86.5%.

2.6. Data dimensionality reduction methods

There are a number of statistical methods, developed over the past few decades, for reducing the dimensionality of data. At the initial stage of the research, the use of the following methods was considered as the most popular and obvious solution to the curse of dimensionality. However, due to the peculiarities and the very nature of the methods described below, their application, despite the validity of their use in solving the problem, would not have the intended effect.

Principal component analysis - PCA - is a statistical technique used to reduce dimensionality in data sets. PCA allows the original variables in a data set to be transformed to a new coordinate system so that the first principal component explains the most variance in the data, the second component explains the next largest portion of the variance, and so on. This process begins with standardizing the data, and then the covariance matrix of the data is calculated. Then, by analyzing the eigenvalues and eigenvectors of the covariance matrix, the principal components are determined. The eigenvectors, called loadings, define the direction of the components in the space of the original variables, while the eigenvalues determine their length, corresponding to the amount of variance explained. Principal components are then used to transform the original variables into a new space, which often allows a significant simplification of the analysis while retaining most of the information contained in the original data [44].

Linear discriminant analysis - LDA - is a technique used in statistics and machine learning that aims to find the linear combination of features that best distinguishes between two or more classes of objects or events. LDA is similar to PCA in that both methods look for linear combinations of variables that are most discriminating. However, while PCA is class-independent and seeks to maximize the variance for all objects, LDA seeks to maximize the difference between classes versus the variance within a class. As a result, LDA is often superior to PCA in cases where the goal is to separate objects belonging to different classes. PCA, on the other hand, may be more effective in cases where the goal is to represent structure in the data without referring to class information [45].

Sammon mapping, also known as Sammon algorithm, is a nonlinear dimensionality reduction technique proposed by John W. Sammon in 1969. Its main goal is to represent multidimensional data in a space with a lower number of dimensions, while minimizing the so-called Sammon mapping error, which is a measure of the difference between the distances of points in the original multidimensional space and the distances of the same points in a space with a lower number of dimensions. This differentiates Sammon mapping from PCA and LDA, which are linear techniques and may not be suitable for mapping complex nonlinear structures. Unlike PCA, Sammon mapping does not aim to maximize variance, and unlike LDA, it does not focus on maximizing separation between classes. Instead, Sammon's mapping aims to preserve the structure of the data by minimizing the deformation of distances between points [46].

However, these methods only serve to change the representation of the data itself by creating combinations of the original features, without changing their input form. When it is important to determine the impact of specific data features on the classification result, it is necessary to use feature relevance determination methods. As with dimensionality reduction methods, these methods are always subject to some error arising from their stochastic or statistical nature.

2.7. Feature selection methods

Data dimensionality reduction and feature selection or extraction are two different categories of techniques used in data processing to improve the efficiency and effectiveness of machine learning models. Dimensionality reduction, as the name suggests, focuses on reducing the number of features in a data set by creating new combinations of features that retain as much relevant information as possible, but in a greatly reduced feature space. Feature selection and extraction, on the other hand, focus on identifying and selecting those features that are most relevant to a given problem. They do not create new features, as in dimensionality reduction, but rather eliminate those features that are less relevant or may introduce noise into the data. By directly translating the results of these methods into data features, in this case individual sensor data, they are ideal for physically selecting the most relevant sensors. These methods can rely on a variety of techniques, including statistical tests, methods based on mutual information or techniques based on machine learning. Several of the methods used in the work are described below.

Minimum Redundancy Maximum Relevance - MRMR - is a feature selection algorithm used in machine learning and bioinformatics. MRMR aims to select the subset of features that are most relevant to the output - maximum relevance - , while minimizing their redundancy with each other - minimum redundancy - . It seeks to do this by simultaneously maximizing the mutual information between features and output and minimizing the sum of mutual information between pairs of features [47].

Boruta, on the other hand, is a feature selection algorithm based on a random forest, proposed by researchers at the University of Warsaw. Boruta seeks to identify all relevant features in a data set by comparing their relevance - measured by their contribution to the random forest model - with the relevance of random features, called "shadows." Features that show greater significance than shadows are considered confirmed, and those that do not are rejected [48]

RelieF is one of the earliest feature selection algorithms, which works by evaluating feature weights based on their ability to distinguish between instances that are close to each other in feature space. In each iteration of the algorithm, RelieF randomly selects an instance from the dataset and then finds its nearest neighbor from the same class - hit example - and its nearest neighbor from a different class - miss example - . The feature weight is updated based on the differences between the selected instance and these two neighbors [49].ReliefF is an improved version of the RelieF algorithm that makes several important improvements. First, ReliefF handles multi-class problems and unbalanced data sets, while the original RelieF algorithm was limited to two-class problems. Second, ReliefF is able to deal with missing values, while RelieF requires complete data [50]

Informational Gain is a feature selection method based on information theory. Informational gain measures how much information about an output class can be obtained by learning the value of a feature. In practice, it is the difference between the entropy of the output class and the conditional entropy of the output class, provided we know the value of the feature. Features with high information gain are considered to be important[51].

Permutational Importance - PI - , also known as permutation-based variable importance, is a technique used to assess the relevance of features in machine learning models such as random forest. PI assesses the relevance of a feature by evaluating how much the accuracy of the model decreases when the values of that feature are randomly permuted. If permutation of feature values leads to a large drop in model accuracy, this suggests that the feature is important to the model's predictions [52].

All of these methods focus on evaluating and selecting significant features from a data set. However, they differ in the way they assess the significance of features and in the specific improvements they make to the feature selection process. However, they are not suitable for assessing the relevance of features in a time series classification problem using neural networks, whether due to the inability to take into account the continuity of time series samples or the need to train a machine learning model in advance. Currently, there are few methods for determining the relevance of features in neural network classification. The most commonly used is the statistical Shapley method.

Shapley values are a concept derived from cooperative game theory that has been adapted to interpret machine learning models. The method assigns each feature a "value" corresponding to its contribution to the model's prediction for a given instance, taking into account all possible combinations of features. A key aspect of the Shapley values is that they are calculated in a "fair" manner, that is, the sum of the Shapley values for all features equals the total model prediction, and the impact of each feature is determined taking into account its interactions with other features [53].

However, in addition to statistical methods, the architects of neural networks are increasingly focusing on ensuring the transparency of the network's prediction by equipping them with elements that facilitate this task.

eXplainable Convolutional Neural Networks - XCNN - is a technique for interpreting the prediction of convolutional networks. XCNN uses special modules called interpretability blocks, which are incorporated into the network architecture to enhance its interpretability. These blocks enable the identification of important areas in the input image that contribute to the model's decisions. In this way, XCNN makes it possible not only to predict the outcome, but also to understand on which areas of the image this prediction was made [54].

3. Design of the data glove

3.1. Choice of sensor technology

Due to the experimental purpose of the constructed glove, it was necessary to select sensors whose properties met the needs of the experimental phase of the device, i.e. durability and simplicity of use in the face of operation. An equally important criterion was the developmental nature of the adopted application. Considering the alternatives - IMU, Hall, optical and mechatronic sensors - , it was decided to use piezoresistive sensors. The technology of piezoresistive sensors is undergoing continuous development, and with the current possibilities, i.e. spraying the sensors directly onto the surface of the glove, thus eliminating the need for joints and related problems, they were considered the most effective method for measuring finger movement [23], [24].

For the construction of the glove described in this paper, piezoresistive sensors from SpectraSymbol were used. The sensors come in two short versions: 73mm and long 110mm. The short sensors have a resistance of 40 kΩ in the upright state and 350 kΩ when bent by 180°, while the long sensors have a resistance of 11 kΩ in the upright state to 38 kΩ when bent by 180°. The values of individual sensors vary slightly within a single batch of product and by as much as 5 to 10 kΩ relative to different batches. In addition, the phenomenon of an increase in the nominal resistance of the sensors by as much as 1 kΩ was noted during the tests.

The manufacturer did not provide the resistance characteristics of the sensors, so it was checked in an experiment. For its purposes, a special stand was created to bend the sensor from 0 to 90°. For the experiment, a 73mm long sensor was used and it was bent in 5° increments at two points 25 and 45mm in length in order to check whether the resistance characteristics depend on the bending point. The resistance was measured using an Axiomet AX-8455 multimeter. The results of the experiment can be seen in Figure 3.1.1. The results of a study of the resistive characteristics of the sensors were confirmed by an independent experiment conducted by another team in [33]

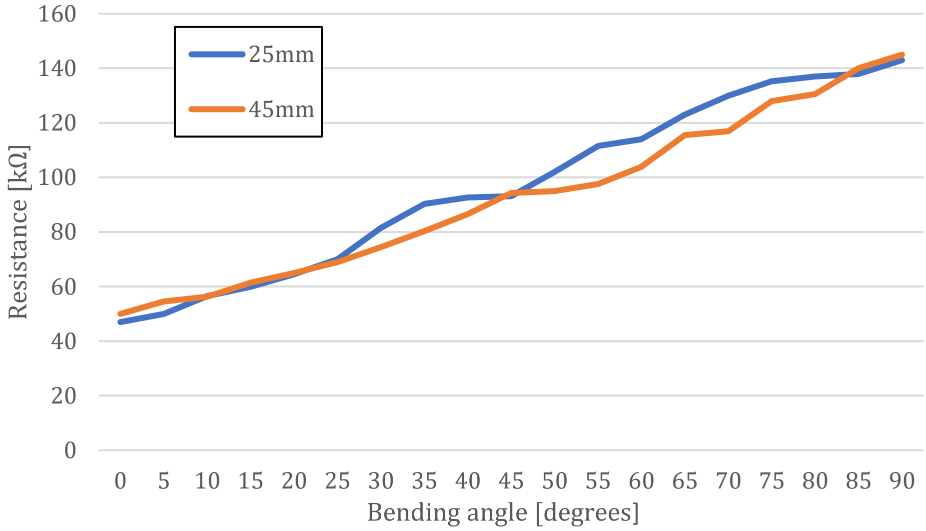

Figure 3.1.1. Measured Resistance characteristics of the sensor

Thus, it can be concluded that the sensors have characteristics close to linear independent of where the sensor is bent. The linear characteristic also means there is no need to convert the values received by the transmitter.

In place of the piezoresistive sensor, the use of other materials was considered: electrically conductive rubber, Velostat or sputtered layer of pure gold. However, none of the mentioned substitutes met the minimum requirements for use. Experiments using rubber showed that its resistive characteristics have hysteresis and take as long as several seconds to return to their original state. Another material considered - Velostat - had virtually zero extensibility, making it unsuitable for the following application. An attempt was also made to sputter a layer of gold onto a layer of the material, but the sample was characterized by the absence of any significant bending response.

The sensors used proved to be an effective and simple-to-use method of tracking finger movement, but not without drawbacks. The sensors have a weak point where the stiffening of the plastic support material of the conductive layer ends. As a result of repeated bending and bending in the opposite direction, they cause cracks to form in the conductive layer, and ultimately to the irreversible destruction of the sensor.

The latest version of the glove also uses the BNO055 sensor. The BNO055 sensor is an 9-axis absolute orientation sensor from Bosch Sensortec. The sensor combines three different sensors: an accelerometer, gyroscope and magnetometer, which together provide information about an object's position, orientation and movement. The BNO055 is also equipped with a microcontroller that performs advanced data processing and calculations on board, providing data directly in the form of Euler angles, quaternions, rotation matrices or linear vectors. The sensor was attached with Velcro to the top of the hand at a location close to its center point.

3.2. Versions of the data glove construction

The design of the glove has undergone three revisions adapting it to the next stage of testing while improving the imperfections of the previous version. The appearance of all three revisions is shown in Figure 3.2.1.

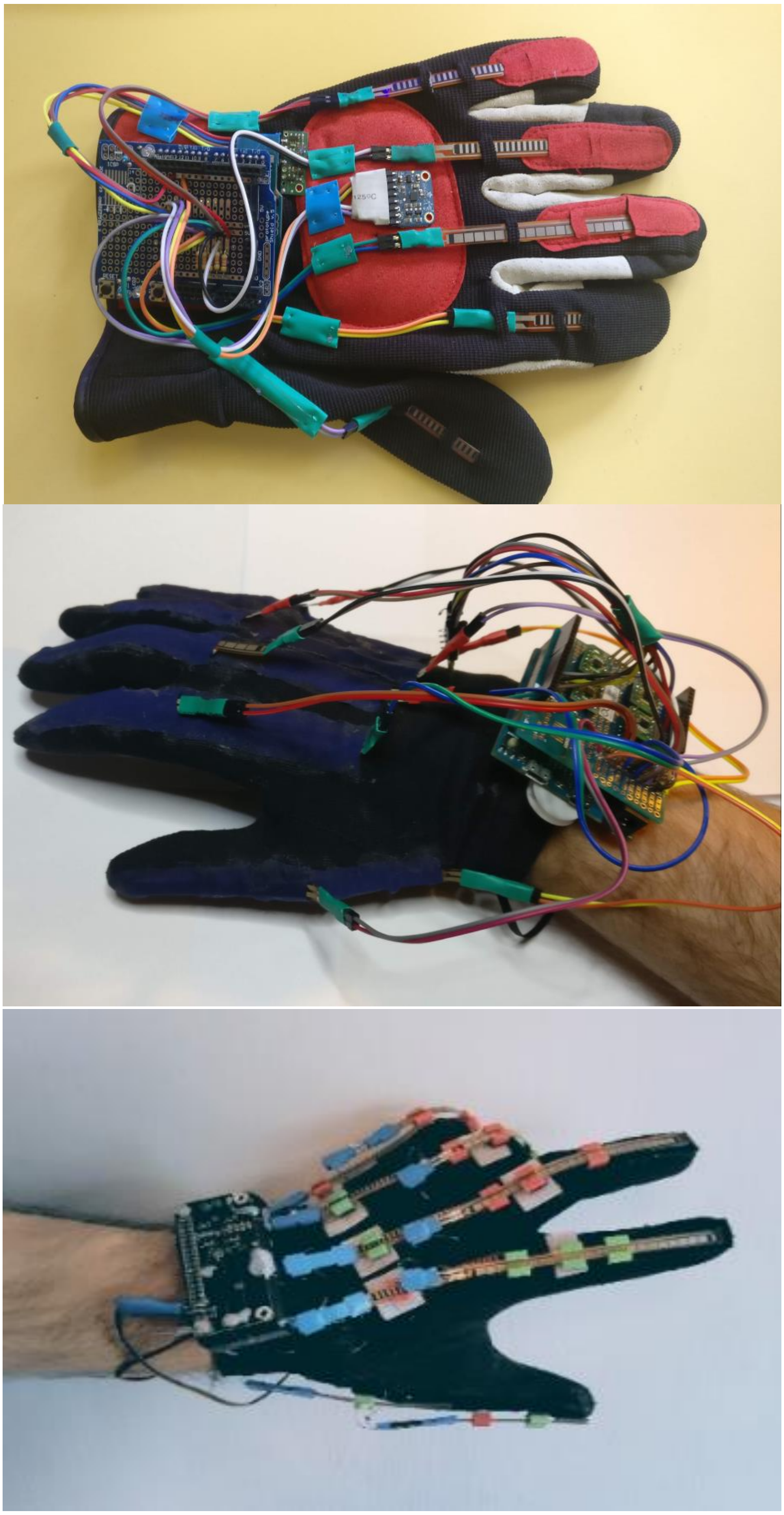

Figure 3.2.1. Photos of three versions of the device

The first version of the glove was designed to evaluate the feasibility of using piezoresistive sensors in the target application. It consisted of 5 sensors: three long ones, placed over the index, middle and ring fingers, and two short ones: on the thumb and little finger.

The sensors were attached to the glove only by sliding them into fabric loops cut out of the glove material. This method of mounting was only sufficient for evaluating the sensors themselves. Mounting them too loosely caused them to slip off the finger during assembly. Connecting the sensors to the measurement and control unit with insulated wires, and mounting these wires with heat shrink sleeves also proved ineffective. The wires were too inflexible and friction against the aforementioned stabilizers from the shirts caused the wires to disconnect from the sensors.

The second revision used a glove with a more snug-fitting design, which allowed the movement of the fingers to be better mimicked by the movement of the sensors. Instead of loops, pockets cut from non-electrostatic synthetic silk bonded to the glove material were used. The number of sensors was also increased to 10. They were placed in two rows: the first over the metacarpophalangeal joints and the second over both interphalangeal joints. The first row consisted solely of 73mm long sensors.

The second row had the same layout as the previous version. Insulated wires were again used as connections, but this time without guides of any kind. This version of the design was used to test the ability to recognize several static gestures based on static PSLA letters. Unsatisfactory results in classifying the gestures tested with it revealed the need to completely change the sensor mounting system.

The current, third version has retained the layout and number of sensors of the previous version, but it has been attached to the glove and connected to the measurement and control unit in a completely new way. The sensors were attached to the surface of the glove using Velcro glued only to the tip of the sensor, while the sensor's motion path was stabilized using 3D printed guides which were also attached to the glove with Velcro. This allowed the sensor to move freely enough to bend with the finger, but stable enough not to move in a messy manner. In addition, the ability to manipulate the position of the sensor allowed their arrangement to be matched to the subject's hand. The use of FCL connectors allowed the sensors to be connected securely and replaced quickly if necessary. Instead of insulated wires, this version used conductive thread which was sewn into the glove material. This allowed connections to be made that did not significantly affect sensor movement and were user-neutral. However, this solution had the disadvantage of creating temporary short circuits between the various signal lines formed by the thread. Attempts to solve this problem using commercially available isolating preparations that would not affect the flexibility of the thread failed. The problem was only mitigated by reorganizing the routing of the individual lines so that they were as far apart from each other as possible and had no slack beyond that necessary for sensor movement.

3.3. Measurement and control unit

In all three iterations of the design, the flex sensors functioned in tandem with a second resistor as a voltage divider whose output signal was passed to the input of the ADC. The converter itself also remained unchanged. The converters used were ADS1115 - 16-bit, four-channel analog-to-digital converters from Texas Instruments, operating at 860 Sps. They communicated with the glove's control unit using the I2C protocol.

With successive stages of research and development of the system, the need for processing power of the control unit increased. The first version used the Arduino Leonardo platform based on an 8-bit ATmega32U4 microcontroller. The next version was built again based on the Arduino platform but this time equipped with a 32-bit ARM Cortex-M0+ ATSAMD21G18 architecture microprocessor. The current version features a Raspberry Pi Zero 2 W single board computer equipped with a Broadcomm BCM2835 processor based on ARM11 architecture, whose operation is controlled by the Linux operating system.

Piezoresistive sensor signal designations are named according to convention:

P[No. of the finger on which the sensor is located]_[No. of the row in which the sensor is located], for example P2.1 is the sensor on the second finger in the first row.

A block diagram of the latest version, along with an illustrative layout and labeling of the various components, is presented in Figure 3.3.1.

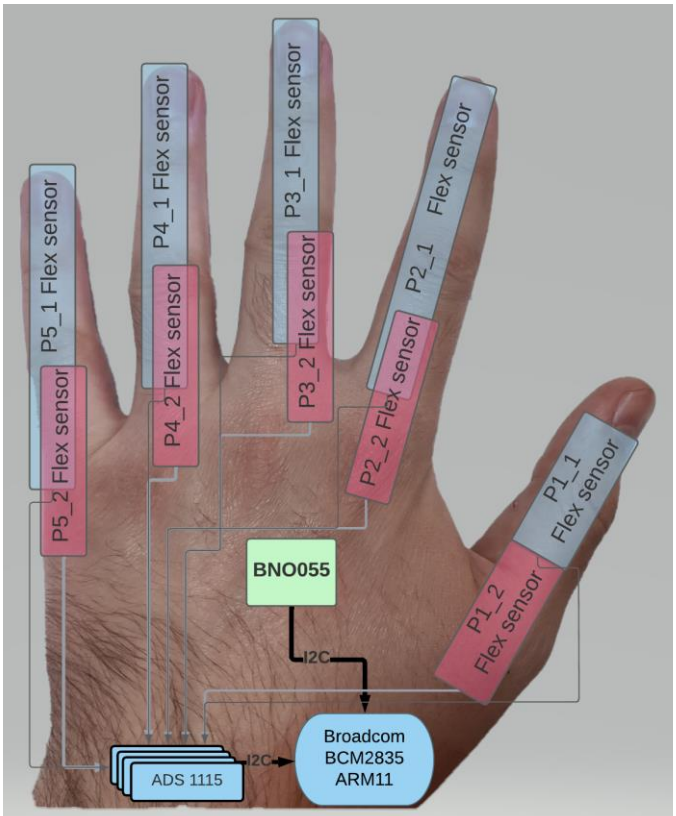

Figure 3.3.1. Block diagram of the device with an illustrative layout of the components

3.4. Glove control system

In the first and second versions, the glove was controlled by wire. The connection was also used for the serial transmission of data from the glove to a computer connected via a serial port. At the stages of testing conducted with these versions of the glove, the wired connection was not too cumbersome for the test subject and did not interfere with the correct performance of the gestures being tested.

In the third version of the glove, due to the intended testing of the correct PSLA letters, a wireless connection via Wi-Fi was used. This was made possible by the SOC used - Raspberry Pi Zero W, which has built-in connectivity of this standard. Thanks to the much greater capabilities of the chip compared to those used in previous versions, the operation of the entire glove was managed by a Linux system whose operation could be controlled via SSH protocol. Initially, both the computer and the glove connected to each other in a hotspot network in which a smartphone played the role of router. However, due to problems with the functioning of such a connection and in order to increase functionality, it was decided to expand the whole system with a separate router and a multi-tasking server. The use of the server made it possible to reduce the use of resources of both the PC and the SOC controlling the glove to a minimum. The server was based on a Raspberry Pi 4 single-board computer, and in the system acted as a Postgres database server, MQTT communication broker and Grafana visualization server.

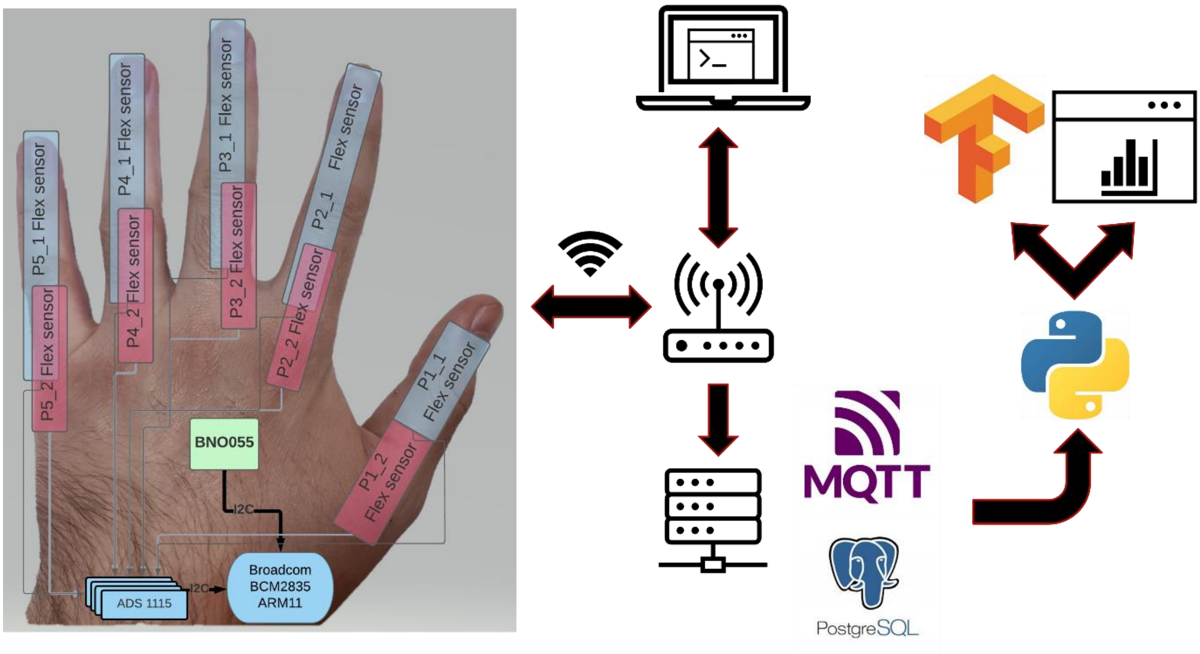

Figure 3.4.1. Symbolic illustration of the system with indicated software and protocols

4. Research methodology

Four stages of research were conducted:

1. Pilot - using the first version of the glove, to test the performance of flex sensors in an application such as a data glove, and to determine the number of sensors and their placement.
2. Evaluation - to test the applicability of the constructed glove in the recognition of static gestures based on PSLA measured from a group of 3 individuals using a second version of the glove.
3. The third stage of the research consisted of a reexamination of static gestures based on the PSLA, however, based on data obtained from a much expanded study group of 15 individuals. The goal of this phase of research was to test the feasibility of reducing the dimension of the features needed to effectively classify the gestures

under study and to determine the features most relevant to this task. The resulting data were classified using the machine learning methods described in Section 4.4. Based on the results obtained, a hierarchy of sensors relevant to the gesture classification process was also determined in order to reduce the dimension of the feature vector.

4. The fourth stage was to study the actual PSLA letters based on data from a study group of similar size - 16 individuals - to the previous one. This stage evaluated the glove's ability to acquire data for recognizing the dynamic hand movements that are PSLAs. Deep learning, described in more detail in Section 4.4, was used for classification at this stage, and several data configuration scenarios for training the neural network were tested to determine the most efficient and effective method.

4.1. Selection of the study group

The pilot studies took place without a research group. During this study, only the performance of the sensors was tested and the design of the glove was evaluated.

Despite the slightly changing research methodology in the next three stages, the criteria for selecting the study group remained the same. We searched for subjects without hand osteo-motor disorders of any age and gender. The key consideration, however, was hand size, measured from the base of the hand at the wrist to the tip of the middle finger. It was intended that the glove would work for people with hand sizes between 16 and 21.5mm measured in the manner described.

During the evaluation study with the second version of the device, three people participated: two women and a man, aged 24, 59 and 66, respectively. One of the test subjects was left-handed. The hand size of the study group members ranged from 18 to 20mm.

In the third part of examination, 15 subjects aged from 23 to 68 volunteered in the experiments. Five women and ten men took part in the trials. Three of the subjects were left-handed. All subjects had an unrestricted range of hand movement necessary for the study. They had similar hand size ranging from 18 to 21 cm measured from the wrist to the tip of the middle finger, with no known musculoskeletal disorders. None of the subjects had also been ex-posed to sign language prior to the study.

The fourth study was conducted on a group of 16 volunteers, men and women between the ages of 21 and 63. Prior to the survey, the size of the hand was measured from the base of the palm to the tip of the index finger. The range of these values was 16.1mm to 20.8mm. None of the test subjects suffered from musculoskeletal disorders of the hand. One of the test subjects is certified in sign language communication.

4.2. Measurement method

Measurements taken at all stages of the study took place under similar conditions. The subject assumed a sitting position with his left arm resting at the elbow on a tabletop or armrest. She would then place her hand in a perpendicular or slightly inclined position to the ground of support. Performing the dynamic letters sometimes required pulling the elbow away from the support to perform a particular letter.

The methodology for obtaining measurements during the evaluation stage consisted of the subject holding a static gesture corresponding to each of the 16 static PSLA letters - shown in Figure 4.2.1. - for about 1s during which a stream of data from all 10 sensors was recorded. At this stage, 48 independent samples were acquired.

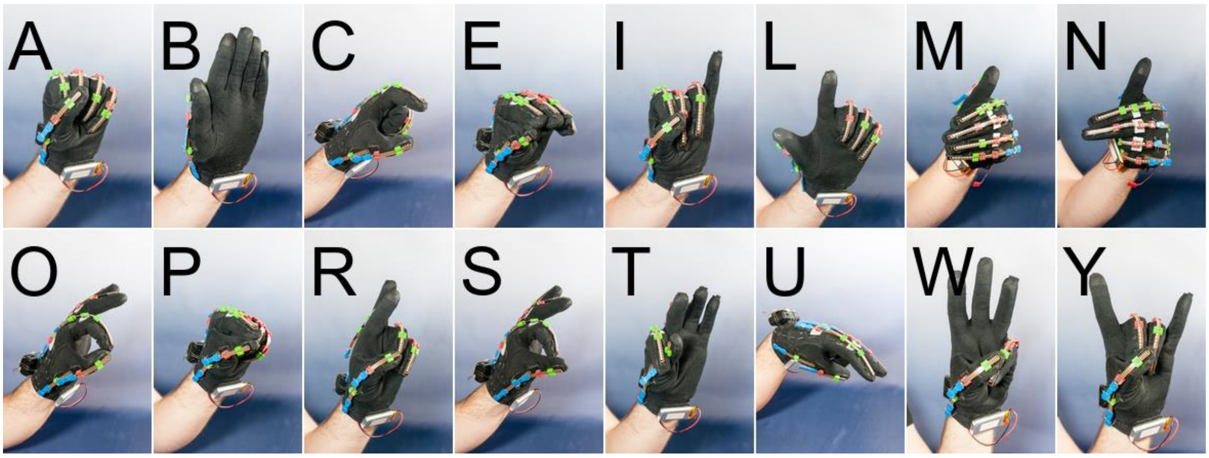

Figure 4.2.1. Static gestures under study

The third stage of the study used a similar data acquisition methodology. Again, each subject was asked to rearrange individual gestures based on static PSLA letters. When a subject performed a given gesture, 10 consecutive vectors containing data from all flex sensors were recorded. About 25,000 independent samples were obtained using this method.

The fourth stage involved collecting measurements as the study participants performed all PSLA letters, both dynamic and static. In order to facilitate data acquisition, the time frame for the execution of each letter was assumed to be, empirically determined, 3 seconds. This allowed for 75 data samples from all 10 piezoresistive sensors and 6 accelerometer axes for each letter of the alphabet. Each participant in the study repeated each letter 10. times. In this round of testing, a total of about 5800 time series were obtained, which corresponds to about 437 thousand individual samples.

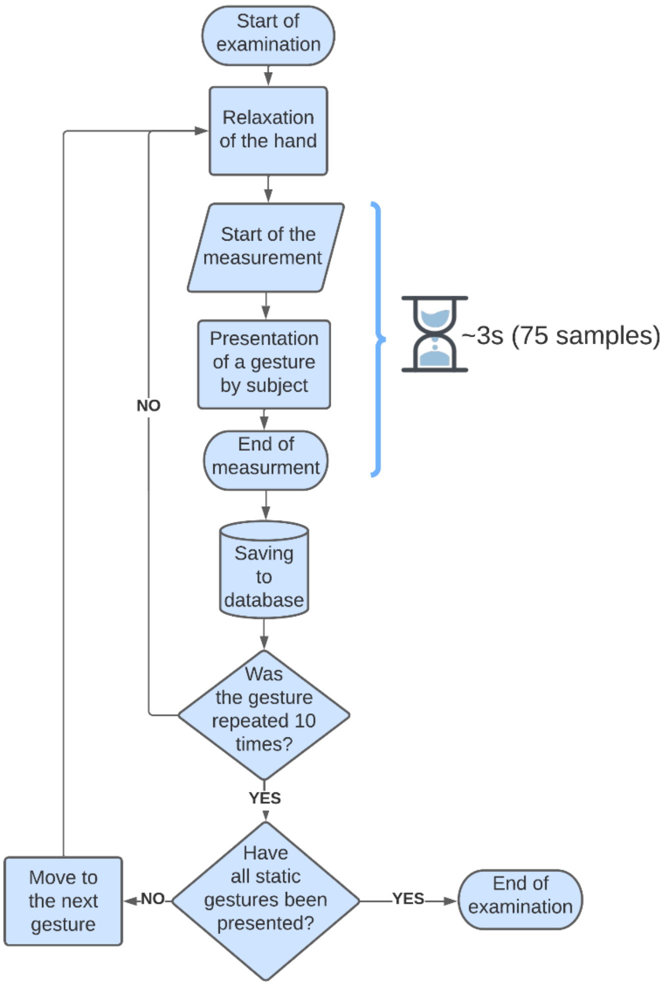

Figure 4.2.2. Block diagram of the methodology of the study regarding the letters of the alphabet of the Polish Sign Language

The tests of the data glove were approved by the Institutional Ethics Committee of Technical University of Lodz - No. 1/2021, date of approval: 28 X 2021 - .

4.3. Data analysis and preprocessing

The processing of the data obtained during the evaluation round of testing consisted largely only of data transformation. This is because the data were recorded in a single column format containing readings from 10 resistive sensors separated by the letter designation of the gesture for which they were collected. In addition, the measurements from each of the three test subjects were placed in separate files. It was therefore necessary to merge the data and convert them into tabular form. Due to the relatively small number of data, this process was carried out manually using Excel.

Data from stage three were stored directly in the memory of the measurement and control unit in the form of CSV files, separate for each subject. They contained data in a tabular format where individual columns corresponded to individual sensors - features - and the categorical designation of the gesture performed. Analysis of the data revealed a significant number of outliers. Short circuits between the power and signal lines of piezoresistive sensors were identified as the cause of their formation. Short circuits during hand movement resulted in a supply voltage at the transducer input, which corresponded to the maximum voltage distinguishable by the transducer of 25,000 in transducer values. It was therefore necessary to select data based on the maximum values of any component in the sample. A value of 22000 in the output values of the transmitter was adopted as the value above which a given sample should be rejected. The merging and selection of data was carried out using Matlab software.

In the case of the fourth stage of the research, thanks to the direct recording of the extracted readings directly into the database, it was not necessary to merge them separately. They were written to the database again in tabular format as in the previous stage, but a column containing time data was added. However, it was necessary to reselect the data due to the short circuits still occurring between the signal and power lines. Thanks to several modifications to the design of the glove, it was possible to significantly reduce the number of them compared to the previous round of tests, and to repeat faulty measurements in real time thanks to data visualization. Selection was carried out on the same basis as in the previous round, with samples rejected if values above 22,000 occurred. Outliers in the form of single, isolated maximum or minimum values also appeared in readings from the inertial sensor. They resulted in a discrepancy in the operation of the chip clock and the SPI communication clock with which the chip was connected to the control unit. Their removal from the obtained waveforms was performed using a moving median filter with a window width of 7 - see Fig. 4.3.1 for illustration of the results - . The same filter was also used on the piezoresistive sensor data to remove noise.

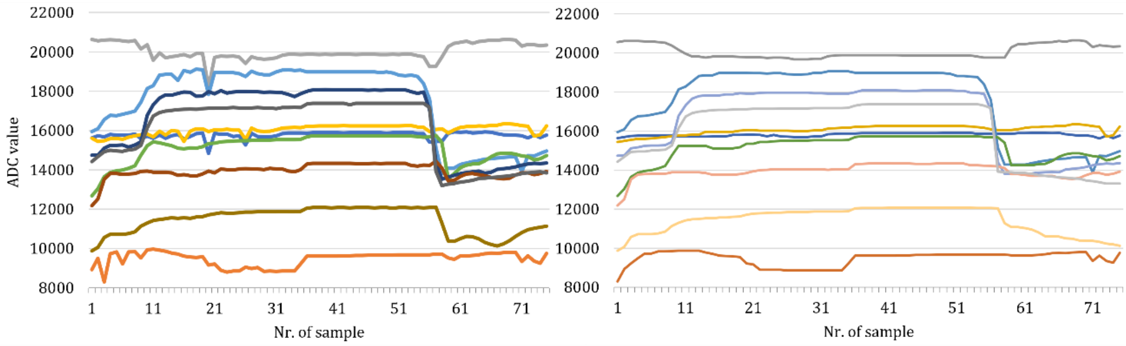

Figure 4.3.1. Visualization of effects of a filter. It can be seen that the noise has been clearly eliminated and that the marginal values of the inertial sensor signals - bottom row - , resulting from hardware defects, have been removed

4.4. Data Augmentation

Using the tsaug package [55], a method was developed to augment the collected database of time series. The augmentation consisted of stretching the data over time for one, two or all three groups: piezoresistive sensors, accelerometer, with a fixed number of 75 samples per single time series. Stretching was random, occurring a maximum of two times per series and lasting no more than 1s total. Noise, which did not exceed 3% of the average signal value, and random drift were also added to the signal. The drift consisted of a linear change in the signal value to a value no greater than 50% of the original signal value over a specified time interval. The results of the signal augmentation can be seen in Figure 4.4.1.

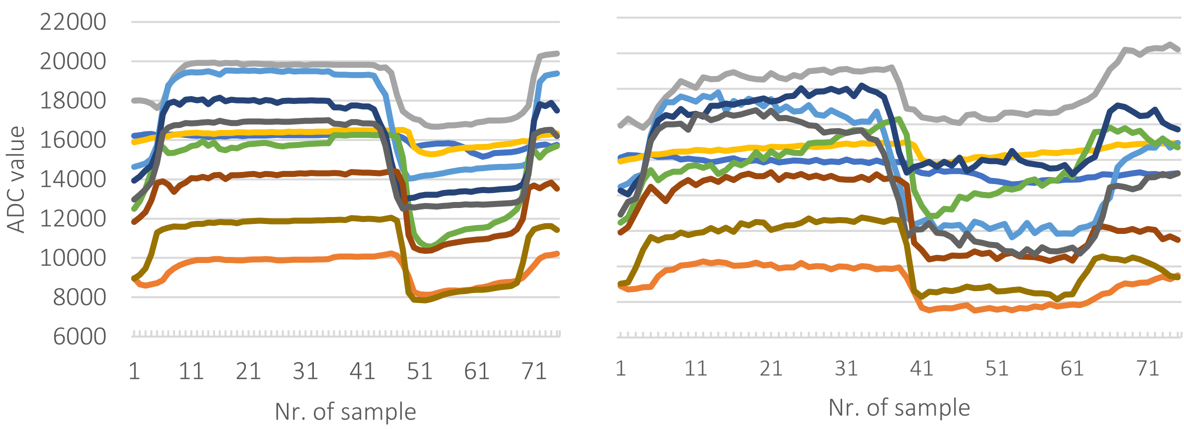

Figure 4.4.1. Illustration of data augmentation. The original signal shown on the left

4.5. Machine learning and deep learning in gestures and PSL Alphabet letter recognition

Machine learning algorithms were used to perform classification of the static hand gestures studied during stages 2 and 3, due to the fact that the acquired data were single, independent samples.

The following classification algorithms were used during stage 2:

- Decision tree
- Naive Bayes classifier
- k-Nearest Neigbors

The third stage involved testing a group of different classifiers, including two composite methods:

- Classification Tree
- Kernel Naive Bayes Classifier
- Cubic SVM
- k-NN
- k-NN in Random Subspaces
- Bagged Trees

Due to the fact that in the final stage, the data are no longer separate samples, but are a time series of 75 related samples. Since the purpose of the study was to determine the possibility of reducing the number of sensors required for proper PSLA letter recognition, no feature extraction methods were used so as not to disturb their original relevance.

The architecture of the network used for classification was chosen based on experiments. Different configurations, the use of LSTMs and GRUs, layer sizes and activation functions were studied.

The neural network model, constructed using the sequential API from the Keras library, consists of multiple layers and activation functions. It accepts input data with a shape corresponding to that of the training data. The first processing step is batch normalization. Then the data passes through a one-dimensional convolution layer - Conv1D - with 64 filters, a kernel of size 3 and a sigmoidal activation function. After another batch normalization, the data is routed to gated recursive units - GRUs - , which use the 'selu' activation function and l2 regularization with a coefficient equal to 0.03. The GRU layer is also equipped with a dropout mechanism, with DP and RDP values for regular and recursive dropout, respectively. This process is performed three times. In the final stage, after batch normalization, the data is routed to a dense layer - Dense - with 36 neurons and a softmax activation function, which is used to classify the results.

Optimization of the model is implemented using the Adam algorithm with an initial learning rate value of LR and a norm that limits the gradient value to 1.

Monitoring of the model training process is done through several feedback mechanisms: TensorBoard for tracking the learning process, recording the best model based on the categorical accuracy on the validation set, reducing the learning rate - ReduceLROnPlateau - when the value of the loss function stops decreasing, and early stopping when the categorical accuracy stops improving.

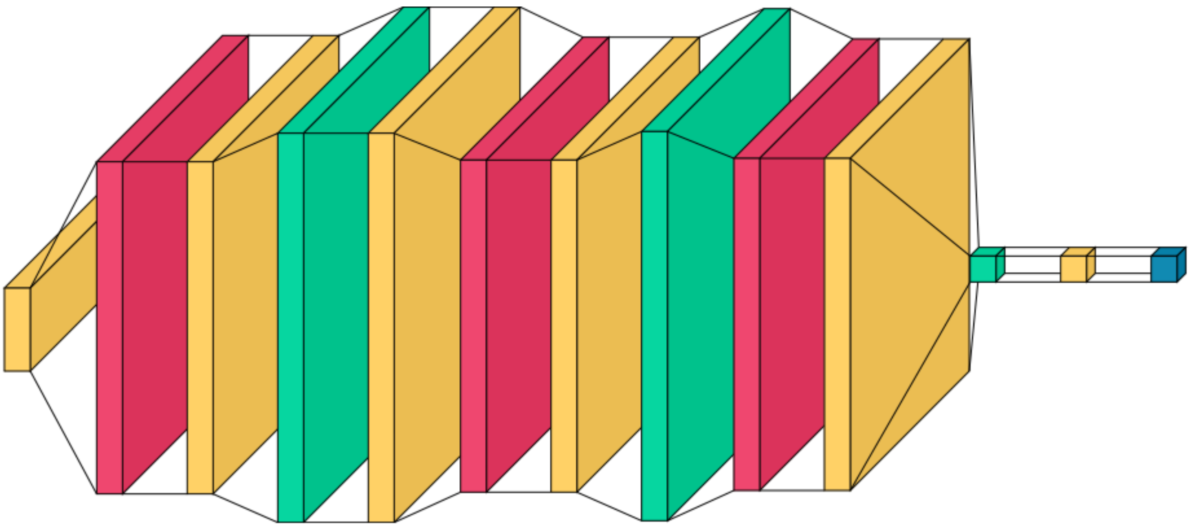

Figure 4.5.1. Neural Network architecture - yellow – Batch Normalization, red – Convolutional Layer, green – GRU Layer, blue - Dense Layer -

4.6. Feature Selection

Initially, feature selection was performed based on the decision tree obtained from the classification process of static gestures based on PSLA letters during the third round of testing. The resulting hierarchy is shown in Figure 4.6.1.

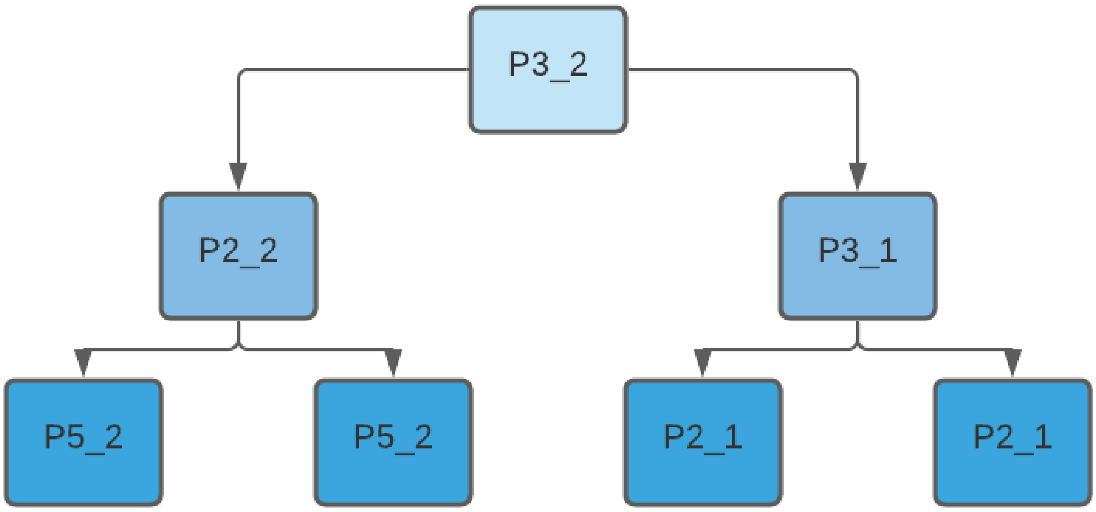

Figure 4.6.1. Peak nodes of the decision tree obtained for the three-component feature vector for static gestures

In addition, the relevance of individual sensors was determined using 4 other methods:

- MRMR
- ReliefF
- Permutational Importance
- Information Gain

The results obtained for each method are shown in Figure 4.6.2.

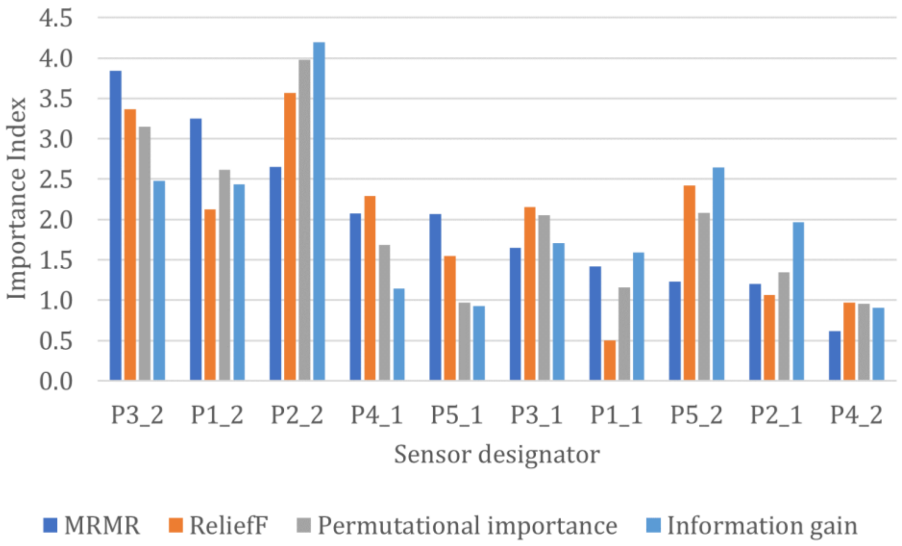

Figure 4.6.2. Importance index results of individual cues obtained using each method

5. Results

The performance of individual classification algorithms was evaluated based on the relevance value expressed by the formula:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \qquad - 1 -$$

Three more indicators of the performance of the classification algorithms were used on a case-by-case basis:

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \qquad - 2 -$$

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \qquad - 3 -$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \qquad - 4 -$$

5.1. Results of preliminary studies

Pilot studies conducted prior to the actual research confirmed the effectiveness of piezoresistive in tracking finger movements of the hand. They also helped determine the optimal placement of the inertial sensor on the hand.

The results of the evaluation study are included in Table 5.1.1. Classification was carried out using five-fold cross-validation. The classification accuracy was expressed using - 1 - .

Table 5.1.1. Effectivness of individual classifiers for the full ten-component feature vector

Classifier	Accuracy
k-NN	41,6%
Naive Bayes Classifier	61,1%
Decision Tree	56,25%

Using the data acquired at this stage, a classification attempt was also carried out on the basis of a reduced feature vector containing readings from sensors over the metacarpophalangeal joints - designation #_2 in Figure 3.3.1 - . The results of this experiment are shown in Table 5.1.2.

Table 5.1.2. Efectiveness of individual classifiers for the three-component feature vector

Classifier	Accuracy
k-NN	50%
Naive Bayes Classifier	66,66%
Decision Tree	41,6%

A detailed description of the research and results was published in [56].

5.2. Results of static gesture classification

The results of stage three on the accuracy of the 6 tested classifiers in recognizing the studied static gestures based on three different sets of features are shown in Figure 5.2.1. The feature vectors were built on the basis of the classification tree obtained by training the algorithm on full-size data.

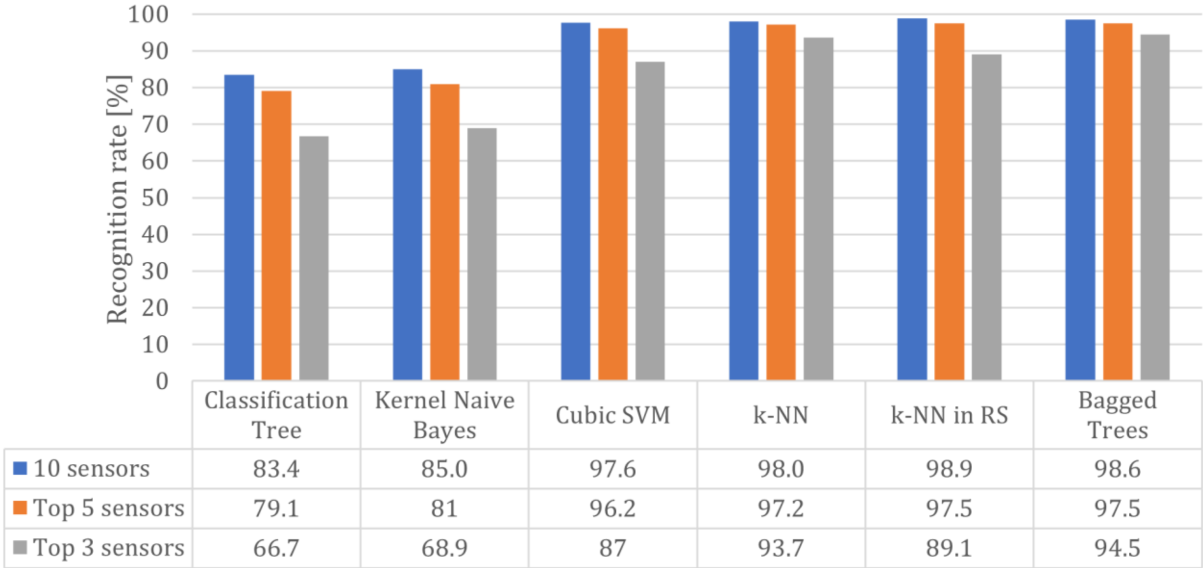

Figure 5.2.1. Effectiveness of individual classifiers for each feature vector size

The average accuracy of each classifier is expressed in Table 5.2.1.

Table 5.2.1. The average accuracy of each classifier

Algorithm	Average Recognition Rate [%]
Classification Tree	76.7
Kernel Naive Bayes Classifier	74.6
Cubic SVM	91.5
k-NN	93.7
k-NN in RS	91.5
Bagged Trees	93.4

The classifiers were repeated in a separate experiment to determine the relevance of sensor position alone. Each of the above six classifiers was re-trained on data containing interphalangeal or metacarpal row data of the sensors - #_1 and #_2 in Figure 3.3.1., respectively - . The results of the classifiers are shown in Figure 5.2.2.

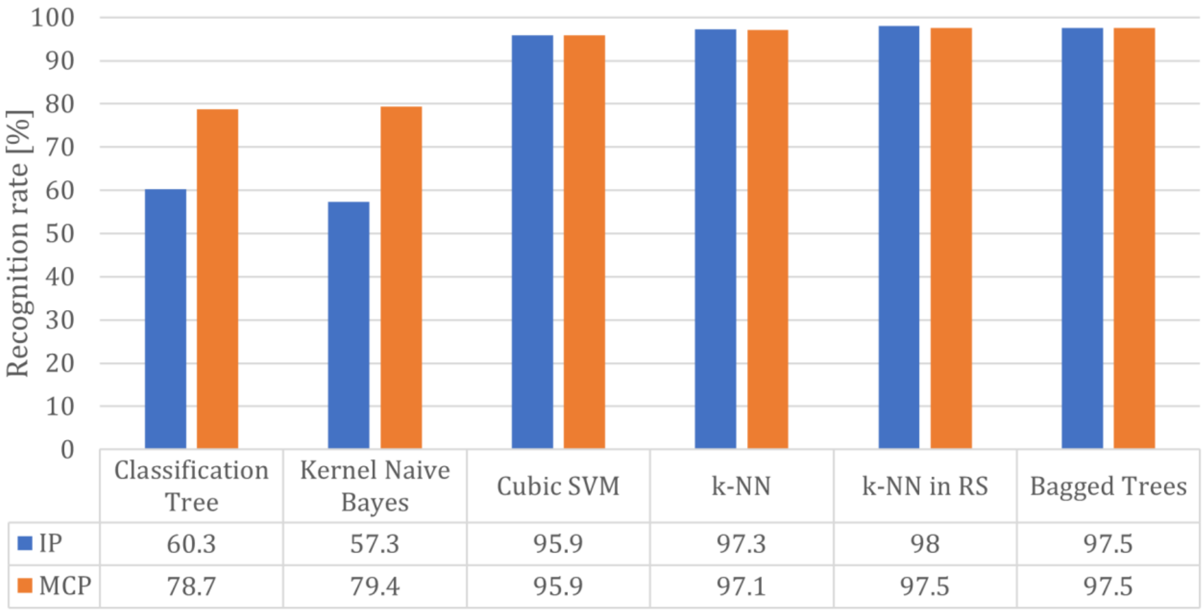

Figure 5.2.2. Effectiveness of the different classifiers according to the distribution over the finger joints - Metacarpophalangeal and Interphalangeal -

Based on the feature hierarchy determined using the feature relevance calculation methods described in Section 4.6, an experiment was conducted to test the classification performance of the top three classifiers based on training on data whose feature vector consisted of 1 to 10 features according to the order of relevance determined by the MRMR method starting with the most relevant feature. The results of the experiment are shown in Figure 5.2.3.

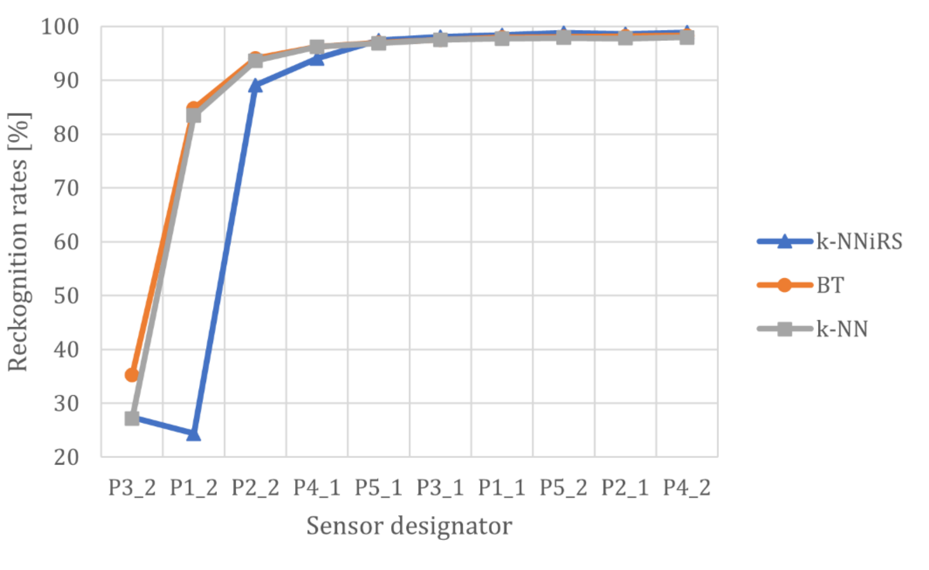

Figure 5.2.3. Effectiveness of individual classifiers when gradually adding readings from individual sensors to the pool of features

The results of this phase of research were published in [57].

Finally, to test the practicality of the methods for determining feature relevance, classification was performed using the k-NNiRS, BT and k-NN algorithms by training them on data with ternary feature vectors consisting of all possible combinations of three of the ten flex sensors. A list of the 5 most effective combinations for each method is shown in Table 5.2.2.

Table 5.2.2. The 5 most effective tercets of features for the three most effective classifiers for static gestures - recurring tercet is shown in blue -

Sensors Designators			**Accuracy**	**Classifier**
P1_1	P3_1	P4_1	92,6%	k-NNiRS
P1_1	P2_2	P4_1	92,4%	
P1_1	P3_2	P4_1	92,2%	
P1_1	P2_1	P4_1	92,0%	
P2_2	P3_1	P4_1	91,6%	
P1_1	P2_1	P4_1	96,9%	BT
P1_1	P3_1	P4_1	96,9%	
P1_1	P1_2	P4_1	96,5%	
P1_1	P1_2	P3_1	96,5%	
P1_1	P3_2	P4_1	96,3%	

P1_1	P3_1	P4_1	95,7%	k-NN
P1_1	P2_1	P3_2	95,4%	
P3_1	P4_1	P4_2	95,3%	
P1_1	P2_1	P4_1	95,2%	
P1_2	P3_1	P4_1	95,1%	

5.3. Results of Polish Sign Language Alphabet letter classification

The next stage of research required the classification of time series, representing individual repetitions of PSLA letters including both static and dynamic hand gestures. First, the most effective neural network learning strategy was tested. Two variants were considered, taking into account the random division into training, validation and test sets in the ratio of 7:2:1:

- Learning from a dataset taken from all subjects
- Learning on a dataset taken from a single subject enriched with data generated using a developed augmenter

100 epochs were provided for training the network in both cases, taking into account early termination of training in case of no improvement in validation accuracy. A mechanism was also used to reduce the learning rate in the absence of improvement in the selected metric. An L2 regularization of $\lambda = 0.03$ and a Dropout of 0.5 were introduced to prevent overlearning of the network. Classification experiments were conducted in parallel on a personal computer and a StandardDS12 computing unit of the Azure platform. The personal computer is equipped with an Intel i9-13900 processor, an RTX 2070 SUPER graphics card and 64GB of RAM. The StandardDS12 is a platform equipped with a quad-core processor and 28GB of RAM.

The network trained on a set consisting of data from all participants reached an accuracy of 76%. The network reached its highest effectiveness around 50 epochs needing about 40 minutes. The result was confirmed by 5-fold cross-validation.

It was decided to select two experiment participants for the data enrichment trial. The criterion for selection was the classification accuracy, measured according to equation no.

- 1 - , achieved for data from one individual when validating the effectiveness of the Leave One Out network. The effectiveness obtained for each individual is shown in Figure 5.3.1.

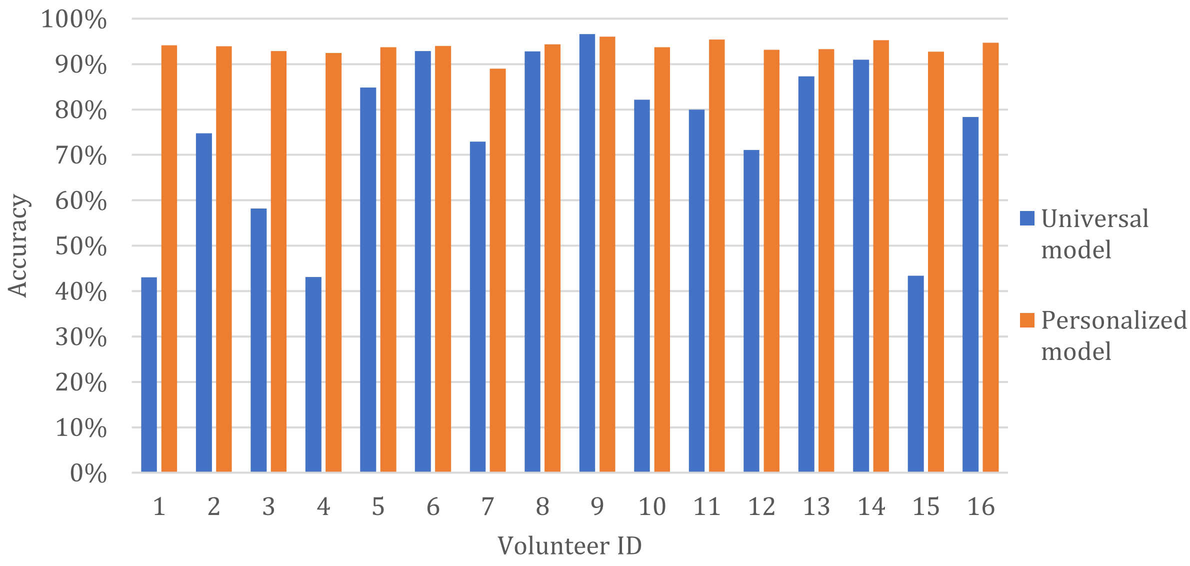

Figure 5.3.1. Classification effectiveness using neural network obtained for individual subjects

The lowest classification score was achieved by data extracted from person no.1 - 42% - , while the highest was from person no.9 - 97% - .

Network training was performed for data from both of the above individuals, with the data divided into training, test, and validation sets, including 5-fold cross-validation. For these two cases, classification using the described network was 89% and 99% accurate, respectively. For a classifier achieving 99% classification accuracy, the sensitivity - 3 - was: 97.6%, specificity - 2 - : 99.9%, and precision - 4 - : 97.7%. The confusion matrix, determined for the case of a classifier with 99% accuracy, can be seen in Figure 5.3.2.

The effectiveness of the developed augmenter was also tested by multiplying 200 times the amount of data from each participant and re-training the model only on data from one person. The classification efficiencies achieved for data from each participant are shown in Figure 5.3.1.

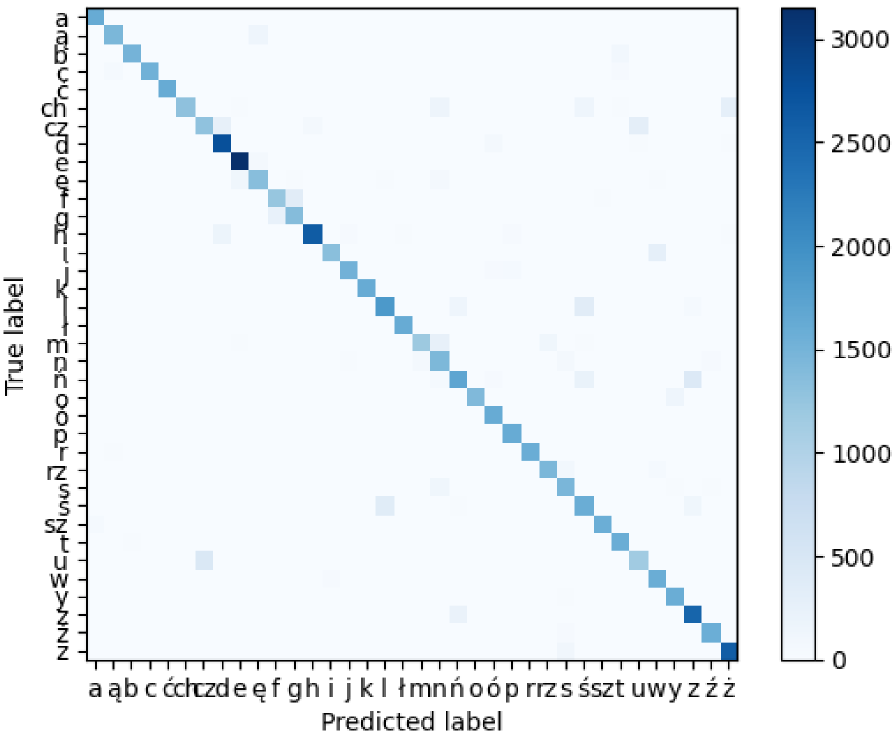

Figure 5.3.2. Confusion matrix obtained for classification of augmented data from a subject #1

To test the collinearity of the data, the Variance Inflation Factor was calculated. The VIF is a statistical measure that is used to identify problems of multicollinearity in regression models where multicollinearity significantly affects results. However, as a measure of multicollinearity, it can also be used in classification models. It is defined by the formula:

$$VIF_i = \frac{1}{1-R_i^2} \qquad \text{- 5 -}$$

Where i denotes a particular variable and R_i denotes coefficient of determination. The coefficient of determination is a key parameter in regression modeling that quantifies the proportion of variation in the dependent variable explained by the independent variables. Mathematically, it is defined as the square of the Pearson correlation coefficient between the observed and predicted values of the dependent variable, reflecting the degree of linear relationship between these values.

Multicollinearity occurs when two or more independent variables in a regression model are highly correlated. This problem can lead to excessively large variance in model parameter estimates, which in turn can lead to model instability and difficulties in interpreting the results. The accepted rule for interpreting VIF results is to consider VIF>10 as an indication of high multicollinearity of an individual feature [58].

Table 5.3.1. Average VIF calculated for each variable for measured dataset

Feature	Average VIF
P1_1	14,7
P1_2	52,3
P2_1	166,6
P2_2	137,7
P3_1	351,0
P3_2	305,0
P4_1	392,1
P4_2	241,6
P5_1	275,0
P5_2	246,8
Euler_x	83,4
Euler_y	84,6
Euler_z	66,3
Acc_x	5,0
Acc_y	4,7
Acc_z	4,7

5.4. Results of Polish Sign Language letter classification based on reduced feature vector

Due to the lack of effective methods for determining the relevance of features in the case of recurrent neural networks, especially those to classify time series data, calculations identical in principle to those for static gestures were carried out. The effectiveness of learning the neural network was tested on multiple data sets containing all possible combinations of three of the ten sensors, as well as data from an inertial sensor. Calculations were also carried out taking into account the second data acquisition scenario - on augmentation data. The obtained accuracies along with the tercets of sensors based on the data from which these effectiveness were obtained are shown in Table 5.4.1.

Table 5.4.1. The 5 most effective tercets of features for Polish Sign Language alphabet classification using neural network

Sensors Designators			Accuracy	Dataset
P1_2	P3_1	P5_1	85,58%	Measured set
P1_2	P3_1	P4_1	84,92%	
P1_2	P2_2	P5_1	84,85%	
P1_2	P4_2	P5_1	84,51%	
P1_2	P3_1	P5_2	84,51%	
P4_2	P5_1	P5_2	99,31%	Augmented set
P2_2	P3_1	P4_2	99,14%	
P1_1	P4_1	P4_2	98,86%	
P1_2	P3_1	P5_1	98,82%	
P1_1	P2_2	P5_2	98,76%	

An analogous experiment was also conducted for a combination of two piezoresistive sensors and for a single piezoresistive sensor. The results of both these experiments are shown in Tables 5.4.2. and on Figure 5.4.1.

Table 5.4.2. The 5 most effective duos of features for Polish Sign Language alphabet classification using neural network

Sensors Designators		Accuracy	Dataset
P1_2	P4_1	82,45%	Measured set
P2_1	P3_1	82,31%	
P2_2	P5_1	81,34%	
P2_1	P3_2	80,40%	
P3_1	P5_2	79,43%	
P2_2	P4_2	99,39%	Augmented set
P4_1	P5_2	98,31%	
P2_2	P3_1	98,31%	
P2_1	P3_2	98,25%	
P2_2	P3_2	98,05%	

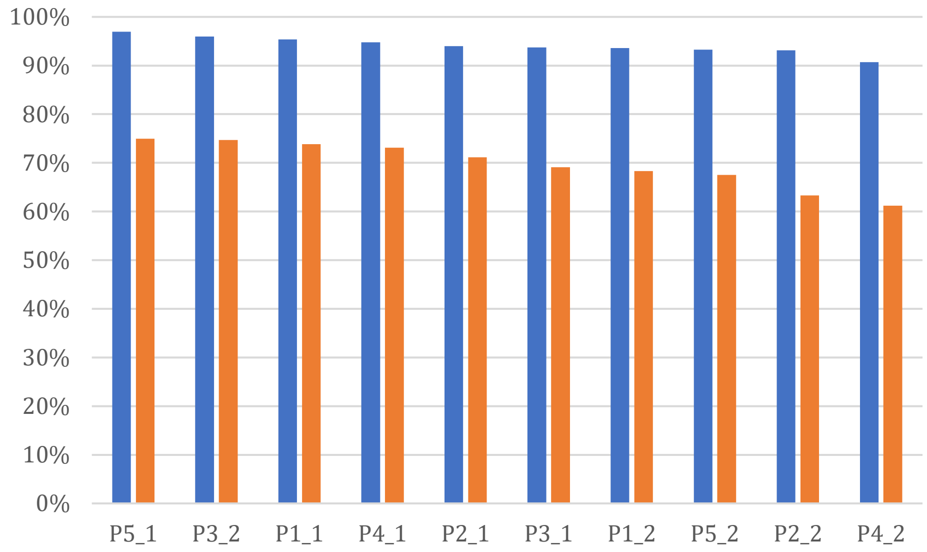

Figure 5.4.1. The effectiveness of particular sensors for Polish Sign Language alphabet classification using neural networks visualized in Figure 4.5.1

5.5. Comparison with existing solutions

It is not possible to directly compare the concept described above with others due to the lack of other proposals for data glove downsizing to recognize hand movements during sign language communication.

Basing on data from [20] where the authors summarized the articles in the field of sensor gloves for hand motion recognition, only 4 articles that dealt with a device that relied on, or functioned properly with, fewer than 5 sensors can be found. However, these articles dealt only with basic gestures or movements. Their list is given in Table 6.3.1.

Table 6.3.1. Articles describing devices with fewer than 5 sensors

No. Ref	# and kind of sensors	Description
[21] Gupta et al. - 2020 -	2 IMU	Forearm movement tracking only
[22] Devnath et al. - 2019 -	2 IMU	Tracking of two fingers only
[23] Zhang et al. - 2019 -	4 pressure sensors	Testing for basic gestures
[24] Huang et al. - 2019 -	2 IMU	Forearm movement tracking only

It is also necessary to compare the effectiveness of the classification algorithms described in this work. Several works on devices with which synthetically separated dynamic gestures or letters of the sign alphabet were studied and classified using neural networks are summarized along with the number of sensors used in this work in Table 6.3.2.

Table 6.3.2. Comparison with results achieved by other studies

No. Ref	# and type of sensors	Highest Accuracy
[17] Simoes Dias et al. - 2022 -	5 Flex + IMU sensors + 2 contact sensors	96,15%
[18] Lee et al. - 2020 -	5 IMU	99%
[19] Pezzuoli et al. - 2021 -	10 Flex + IMU sensors	99,7%
[16] Bae et al. - 2021 -	10 Flex + IMU sensors	94%
[20] Saggio et al. - 2020 -	10 Flex + IMU sensors	96,6%
This Work	3 Flex + IMU	99%

Thus, as can be seen, the concept described in this paper makes it possible to achieve equally good if not better results using a number of sensors even several times smaller than comparable solutions.

6. Discussion and Conclusions

6.1. Summary of research results

The initial pilot stage of the research allowed positive verification of the overall performance of the first version of the glove. The conclusions drawn from these results guided the work on the next version of the glove.

The second version, already consisting of 10 sensors, was used to carry out measurements culminating in an attempt to classify the obtained data using basic machine-learning algorithms during the evaluation stage of the research. However, the low classification performance indicated the need to improve a number of aspects of the glove, both structural and related to the software and the method of performing the measurements themselves.

A third version of the glove was constructed to make the corrections necessary for the glove to function properly. Using it, the final two stages of testing were carried out: testing of static gestures and full-fledged PSLs.

The results of the third stage of the study showed a slight decrease in the classification performance of the decision tree, and even an increase in the classification performance of the other two algorithms using only a three-component feature vector for the classification process, as shown in Tables 5.4.1. and 5.4.2. This implied that similar results could be achieved in the classification of more complex gestures and ultimately in the classification of PSLA letters. This conclusion was also supported by the observed need to improve the acquisition process and the design itself, as it was only assumed to increase the recognition efficiency of the gestures under study using the full ten-component feature vector

The results obtained in the PSLA letter-based static gesture classification experiment confirmed this conjecture. As Figure 4.6.2. and Table 5.2.1. show, limiting the size of the feature vector to only three elements in only a slight reduction in relevance for most of the algorithms tested. In order to specify which sensors are the most relevant from the point of view of classification performance, several methods were used to determine the relevance of the features, the results of which are shown in Figure 4.6.2. Considering also the individual nodes of the decision tree obtained in the classification process, it was found that from the point of view of classification of static gestures based on PSLA letters - shown in Figure 4.2.1 - , the most relevant sensors are, in turn: 3_2, 1_2 and 2_2. However, an experiment involving classification using all possible combinations of the three features, the results of which are shown in Tab 5.4.2, showed markedly different results. According to this experiment, the most significant sensors are: 1_1, 3_1, 4_1. The discrepancy between the significance determined by statistical methods and that proved by experiment is to be found in the multicollinearity of the obtained data, shown by the high VIF index [59]. Multicollinearity results in low reliability of statistical methods, so the following experiments assumed the occurrence of a similar phenomenon and the inability to effectively estimate the significance of individual characteristics.

An experiment on the recognition of all PSLA letters was launched to evaluate potential neural network training scenarios for hypothetical end-user adaptation. The results of this study demonstrate the effectiveness of the data augmenter and its use in the calibration process. The network trained on augmented data from only one subject at the time achieved improved accuracy for all subjects. As high as 99% classification accuracy for data from subject #9 and 94% from subject #1, improving the accuracy of classifications made by this volunteer by more than double that of the network trained on data from all subjects.

Among the highest results achieved for the set of all letters and the set of augmented readings from one person, highlighted in Table 5.2.2., was the tercet: 1_2, 3_1, 5_1 - marked in blue - . Among the results of the experiment on the combination of two sensors, it is also possible to pick out a repeating duo for both data sets: P2_1, P3_2. More importantly, however, the achieved classification performance on the basis of data from only two piezoresistive sensors and a 6-axis inertial sensor allows achieving a classification accuracy of 99% on the augmented data and 82% for the data from all test subjects combined =. This demonstrates another advantage of using an approach that roots out data from only one person to train the network.

The results of calculating the classification effectiveness of the proposed neural network using data from only one sensor, shown in Figure 5.4.1, also testify in favor of the thesis of this paper. For, depending on the data set, the accuracy reached up to 96%. In both cases, too, the feature allowing the highest accuracy was the data from sensor 5_1, being part of the tercet appearing among the 5 tercets resulting in the best classification accuracy for both data sets.

By proving the thesis of this work, i.e. the possibility of effective recognition of gestures and, in particular, letters of the Polish Sign Language alphabet, using only three sensors, the possibility of reducing the size of gloves for recognizing letters of the Polish Sign Language alphabet has been proven. This is extremely important for the prospect of developing such gloves as part of a sign language interpreting system, as lack of ergonomics and low efficiency are common criticisms of this technology. So far, none of the teams conducting research and similar devices has carried out experiments to determine the relevance of individual sensors, on the assumption that it is necessary to have at least 5 of them.

6.2. Contribution and practical implications

Summarizing the achieved results:

- It has been proven that it is possible to successfully recognize synthetically segmented letters of Polish Sign Language using data from only three or even a single piezoresistive sensor with an accuracy not worse than 90%.

- The advantage of the strategy of creating personalized models for the recognition of PSL letters and universal model was shown.
- An effective method of augmentation of the obtained data was presented, allowing to reduce the time required to acquire the data necessary to create an effective model.

The results of the following work can be used by designers and researchers to direct their technology towards reducing the size of gloves that are part of hypothetical sign language translation systems.

Identified potential uses for sign alphabet letter recognition gloves, which could prove particularly beneficial for sign language-only users, autistic people and in emergency situations. While these gloves are not a full-fledged sign language interpreter, they could be a valuable part of a communication support system.

The most important achievement is to prove the possibility of reducing the size of such gloves by reducing the number of sensors. As part of our work, a neural network was trained using only data from three finger flex sensors and data from a six-axis inertial sensor. The study shows that such minimization of sensors not only does not reduce the effectiveness of gesture recognition, but can also help reduce production costs and increase the ergonomics of the device.

The fact of the appearance of identical combinations of three features in the best results during the brute-force feature selection process suggests that there is a universal and effective combination of features that can be applied to different people, and that can be represented by piezoresistive sensors placed in specific locations on the hand.

Intriguing results were also produced by a study in which, by augmenting data from a single person, it was possible to achieve higher gesture recognition performance than when training on a set of data from multiple people. This phenomenon indicates that extensive measurements from a large number of people are not necessarily required to create a universal model. A short calibration performed on a single person appears to be sufficient, greatly simplifying the process of creating effective gesture recognition systems.

In addition, these findings pave the way for the creation of a global, universal model through the use of transfer learning. This method involves using previously learned neural network models and retraining them on new data, saving computational resources and time. By using data from multiple people, it is possible to adapt the model to different styles and idiosyncrasies of hand movements, which can contribute to a more versatile and universal gesture recognition system.

However, while the results are promising, these applications are hypothetical for now and require further in-depth research. It is not yet clear to what extent this technology can help improve communication in emergency situations, or how it will affect the comfort and quality of life of sign language users or autistic people. Moreover, it is also important to conduct further research to fully understand how reducing the size and number of sensors will affect the long-term performance and durability of such a device. Nonetheless, the findings open up new possibilities for future innovations in the field of assistive communication technology.

6.3. Analysis of potential constraints and challenges

Despite the ability to effectively recognize gestures and letters of the sign alphabet, the glove alone is not a full-fledged and sufficient tool for sign language translation on its own. It must therefore be part of a broader system that would recognize the other elements of sign language syntax. In its current form, the device is not designed to recognize hand movements in real time. It is necessary to develop an algorithm that effectively segments the data from all the sensors of the glove. This is a highly complex challenge due to the unregulated and individual style of SL use by each user.

Development can also occur in the area of sensor technology. The piezoresistive flex sensors used in the glove described above could be replaced by a piezoresistive layer sprayed directly onto the glove material. This would allow a reduction in connection-related failures and in the glove dimensions themselves.

6.4. Prospects for further research

As the next stage of research, it is envisaged to modify the device to allow the study of gestures and letters performed in a string, without their synthetic separation. It will therefore be necessary to develop or implement an existing segmentation algorithm. An intermediate step to achieve this goal will be to use the readings currently obtained by trimming them for redundant hand resting states.

In addition to the limitations mentioned above requiring further research and development of the device itself, there are also worthwhile prospects for using the device in other fields. The glove could find application in rehabilitation as a diagnostic tool for detecting dysfunctions of the hand's musculoskeletal system and the conditions causing them. It would also be possible to assess the progress of rehabilitation using readings obtained with the glove.

Another area in which the glove could find application is telecontrol. With the increasing use of drones in many, often dangerous, tasks comes the need for anthropomorphic grippers. Since the grasps performed by the human hand can be interpreted as gestures, it is not out of the question that a simplification of the design of the data glove can also be implanted to control such manipulators.

7. References

[1] “Assistive listening devices / Deaf and Hard of Hearing.” https://mn.gov/deaf-hard-of-hearing/assistive-technology/assistive-listening-devices/ - accessed Jul. 14, 2023 - .

[2] “Assistive Devices for People with Hearing or Speech Disorders | NIDCD.” https://www.nidcd.nih.gov/health/assistive-devices-people-hearing-voice-speech-or-language-disorders#4 - accessed Jul. 14, 2023 - .

[3] “Tłumacz Migam - Migam.” https://migam.org/tlumacz-migam/ - accessed Jul. 14, 2023 - .

[4] “Virtual Reality - VR - Market Size & Share Analysis - Industry Research Report - Growth Trends.” https://www.mordorintelligence.com/industry-reports/virtual-reality-market - accessed Jul. 14, 2023 - .

[5] “The Effects Of Virtual Reality - VR - On Society | Bernard Marr.” https://bernardmarr.com/the-effects-of-virtual-reality-vr-on-society/ - accessed Jul. 14, 2023 - .

[6] “6 Best VR Controllers 2023 For All VR Players - Tech10.” https://techonexero.com/best-vr-controllers/ - accessed Jul. 14, 2023 - .

[7] Emily A. Vogels, “Millennials stand out for their technology use | Pew Research Center,” Aug. 09, 2019. https://www.pewresearch.org/fact-tank/2019/09/09/us-generations-technology-use/ - accessed Feb. 13, 2021 - .

[8] “The Somatosensory System | Boundless Anatomy and Physiology.” https://courses.lumenlearning.com/boundless-ap/chapter/the-somatosensory-system/ - accessed Feb. 13, 2021 - .

[9] “Approach - Neuralink.” https://neuralink.com/approach/ - accessed Feb. 16, 2021 - .

[10] G. ElKoura and K. Singh, “Handrix: Animating the human hand,” in *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA 2003*, 2003.

[11] G. S. Gandhoke, E. Belykh, X. Zhao, R. Leblanc, and M. C. Preul, “Edwin Boldrey and Wilder Penfield’s Homunculus: A Life Given by Mrs. Cantlie - In and Out of Realism - ,” *World Neurosurgery*, vol. 132. Elsevier Inc., pp. 377–388, Dec. 01, 2019. doi: 10.1016/j.wneu.2019.08.116.

[12] D. J. Sandin and T. A. DeFanti, “Sayre Glove - US NEA R60-34-163 Final Project Report.” pp. 1–30, 1977.

[13] “Assistive Technology for Deaf and Hard of Hearing Students | Verbit.” https://verbit.ai/assistive-technology-for-deaf-and-hard-of-hearing-students/ - accessed Jul. 14, 2023 - .

[14] “Why the Signing Gloves Hype Needs to Stop - Audio Accessibility.” https://audio-accessibility.com/news/2016/05/signing-gloves-hype-needs-stop/ - accessed Dec. 14, 2021 - .

[15] A. Grieve-Smith, “Ten reasons why sign-to-speech is not going to be practical any time soon. – Technology and language.” http://grieve-smith.com/blog/2016/04/ten-reasons-why-sign-to-speech-is-not-going-to-be-practical-any-time-soon/ - accessed Dec. 14, 2021 - .

[16] A. Wadhawan and P. Kumar, “Sign Language Recognition Systems: A Decade Systematic Literature Review,” *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 785–813, 2021, doi: 10.1007/s11831-019-09384-2.

[17] R. Rastgoo, K. Kiani, and S. Escalera, “Sign Language Recognition: A Deep Survey,” *Expert Syst Appl*, vol. 164, no. July 2020, p. 113794, 2021, doi: 10.1016/j.eswa.2020.113794.

[18] N. T. Do, S. H. Kim, H. J. Yang, and G. S. Lee, “Robust hand shape features for dynamic hand gesture recognition using multi-level feature LSTM,” *Applied Sciences - Switzerland -* , vol. 10, no. 18, Sep. 2020, doi: 10.3390/APP10186293.

[19] “Technology - VRfree® glove - intuitive VR interaction.” https://www.sensoryx.com/company/technology/ - accessed Feb. 07, 2021 - .

[20] Z. R. Saeed, Z. B. Zainol, B. B. Zaidan, and A. H. Alamoodi, “A Systematic Review on Systems-Based Sensory Gloves for Sign Language Pattern Recognition: An Update from 2017 to 2022,” *IEEE Access*, pp. 1–1, 2022, doi: 10.1109/ACCESS.2022.3219430.

[21] M. Rinalduzzi *et al.*, “Gesture recognition of sign language alphabet using a magnetic positioning system,” *Applied Sciences - Switzerland* - , vol. 11, no. 12, Jun. 2021, doi: 10.3390/app11125594.

[22] “LucidVR/lucidgloves: Arduino/ESP32 based DIY VR Haptic gloves. Compatible with SteamVR via OpenGloves.” https://github.com/LucidVR/lucidgloves - accessed Apr. 30, 2023 - .

[23] K. Aw, J. Budd, and T. Wilshaw-Sparkes, “Data Glove Using Soft and Stretchable Piezoresistive Sensors,” *Micromachines - Basel* - , vol. 13, no. 3, 2022, doi: 10.3390/mi13030372.

[24] M. Lee and J. Bae, “Deep Learning Based Real-Time Recognition of Dynamic Finger Gestures Using a Data Glove,” *IEEE Access*, vol. 8, pp. 219923–219933, 2020, doi: 10.1109/ACCESS.2020.3039401.

[25] M. Lee and J. Bae, “Real-Time Gesture Recognition in the View of Repeating Characteristics of Sign Languages,” *IEEE Trans Industr Inform*, vol. 18, no. 12, pp. 8818–8828, 2022, doi: 10.1109/TII.2022.3152214.

[26] Z. Zhang, “Introduction to machine learning: K-nearest neighbors,” *Ann Transl Med*, vol. 4, no. 11, pp. 1–7, 2016, doi: 10.21037/atm.2016.03.37.

[27] T. K. Ho, “Nearest neighbors in random subspaces,” *Lecture Notes in Computer Science - including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics* - , vol. 1451, pp. 640–648, 1998, doi: 10.1007/bfb0033288.

[28] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach Learn*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[29] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989, doi: 10.1109/5.18626.

[30] O. Gold and M. Sharir, “Dynamic Time Warping and Geometric Edit Distance,” *ACM Transactions on Algorithms*, vol. 14, no. 4, pp. 1–17, Oct. 2018, doi: 10.1145/3230734.

[31] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[32] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," Dec. 2014, doi: https://doi.org/10.48550/arXiv.1412.3555.

[33] T. Simoes Dias, J. J. A. M. Junior, and S. F. Pichorim, "An Instrumented Glove for Recognition of Brazilian Sign Language Alphabet," *IEEE Sens J*, vol. 22, no. 3, pp. 2518–2529, 2022, doi: 10.1109/JSEN.2021.3136790.

[34] B. G. Lee, T. W. Chong, and W. Y. Chung, "Sensor Fusion of Motion-Based Sign Language Interpretation with Deep Learning," *Sensors 2020, Vol. 20, Page 6256*, vol. 20, no. 21, p. 6256, Nov. 2020, doi: 10.3390/S20216256.

[35] F. Pezzuoli, D. Corona, and M. L. Corradini, "Recognition and Classification of Dynamic Hand Gestures by a Wearable Data-Glove," *SN Comput Sci*, vol. 2, no. 1, 2021, doi: 10.1007/s42979-020-00396-5.

[36] G. Saggio, P. Cavallo, M. Ricci, V. Errico, J. Zea, and M. E. Benalcázar, "Sign language recognition using wearable electronics: Implementing K-nearest neighbors with dynamic time warping and convolutional neural network algorithms," *Sensors - Switzerland -* , vol. 20, no. 14, pp. 1–14, 2020, doi: 10.3390/s20143879.

[37] R. Gupta and S. Rajan, "Comparative Analysis of Convolution Neural Network Models for Continuous Indian Sign Language Classification," *Procedia Comput Sci*, vol. 171, pp. 1542–1550, Jan. 2020, doi: 10.1016/J.PROCS.2020.04.165.

[38] P. P. Devnath, A. S. Kundu, O. Mazumder, and S. Bhaumik, "Human—Hand Posture Classification For Robotic Teleoperation Using Wearable Sensor," in *2019 IEEE Region 10 Symposium - TENSYMP -* , IEEE, Jun. 2019, pp. 114–119. doi: 10.1109/TENSYMP46218.2019.8971080.

[39] Y. Zhang, B. Liu, and Z. Liu, "Recognizing Hand Gestures With Pressure-Sensor-Based Motion Sensing," *IEEE Trans Biomed Circuits Syst*, vol. 13, no. 6, pp. 1425–1436, 2019, doi: 10.1109/TBCAS.2019.2940030.

[40] X. Huang *et al.*, "Tracing the Motion of Finger Joints for Gesture Recognition via Sewing RGO-Coated Fibers onto a Textile Glove," *IEEE Sens J*, vol. 19, no. 20, pp. 9504–9511, 2019, doi: 10.1109/JSEN.2019.2924797.

[41] P. Pławiak, T. Sośnicki, M. Niedźwiecki, Z. Tabor, and K. Rzecki, "Hand Body Language Gesture Recognition Based on Signals From Specialized Glove and Machine Learning Algorithms," *IEEE Trans Industr Inform*, vol. 12, no. 3, pp. 1104–1113, 2016, doi: 10.1109/TII.2016.2550528.

[42] T. Dziubich and J. Szymański, "Automatic classification of Polish sign language words | Automatyczna klasyfikacja znaków Polskiego je{ogonek}zyka miganego," *Przeglad Elektrotechniczny*, vol. 90, no. 3, 2014, doi: 10.12915/pe.2014.03.43.

[43] E. Korzeniewska, M. Kania, and R. Zawiślak, "Textronic Glove Translating Polish Sign Language," *Sensors*, vol. 22, no. 18, 2022, doi: 10.3390/s22186788.

[44] *Principal Component Analysis*. New York: Springer-Verlag, 2002. doi: 10.1007/b98835.

[45] A. M. Martinez and A. C. Kak, “PCA versus LDA,” *IEEE Trans Pattern Anal Mach Intell*, vol. 23, no. 2, pp. 228–233, 2001, doi: 10.1109/34.908974.

[46] J. W. Sammon, “A Nonlinear Mapping for Data Structure Analysis,” *IEEE Transactions on Computers*, vol. C–18, no. 5, pp. 401–409, May 1969, doi: 10.1109/T-C.1969.222678.

[47] Hanchuan Peng, Fuhui Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Trans Pattern Anal Mach Intell*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.

[48] M. B. Kursa and W. R. Rudnicki, “Feature Selection with the **Boruta** Package,” *J Stat Softw*, vol. 36, no. 11, 2010, doi: 10.18637/jss.v036.i11.

[49] I. Kononenko, “Estimating attributes: Analysis and extensions of RELIEF,” 1994, pp. 171–182. doi: 10.1007/3-540-57868-4_57.

[50] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, “Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF,” *Applied Intelligence*, vol. 7, no. 1, pp. 39–55, 1997, doi: 10.1023/A:1008280620621.

[51] J. R. Quinlan, “Induction of decision trees,” *Mach Learn*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.

[52] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, “Permutation importance: a corrected feature importance measure,” *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, May 2010, doi: 10.1093/bioinformatics/btq134.

[53] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[54] R. Ibrahim and M. O. Shafiq, “Explainable Convolutional Neural Networks: A Taxonomy, Review, and Future Directions,” *ACM Comput Surv*, vol. 55, no. 10, pp. 1–37, Oct. 2023, doi: 10.1145/3563691.

[55] T. Wen and R. Keyes, “Time Series Anomaly Detection Using Convolutional Neural Networks and Transfer Learning,” May 2019, doi: 10.48550/arxiv.1905.13628.

[56] J. PISKOZUB, “Rozpoznawanie gestów statycznych ręki za pomocą rękawicy sensorowej,” *PRZEGLĄD ELEKTROTECHNICZNY*, vol. 1, no. 9, pp. 144–147, Sep. 2019, doi: 10.15199/48.2019.09.29.

[57] J. Piskozub and P. Strumillo, “Reducing the Number of Sensors in the Data Glove for Recognition of Static Hand Gestures,” *Applied Sciences - Switzerland* - , vol. 12, no. 15, 2022, doi: 10.3390/app12157388.

[58] J. Lambrinos, J. Neter, W. Wasserman, and M. H. Kutner, “Applied Linear Regression Models,” *Technometrics*, vol. 26, no. 4, p. 415, Nov. 1984, doi: 10.2307/1269508.

[59] R. Snee, “Who Invented the Variance Inflation Factor?” 1981. doi: 10.13140/RG.2.1.3274.8562.

8. Supplement

8.1. Acquired data, codes and the course of the experiments

The data obtained during the experiments, all Python scripts used during experiments and the templates of the form informing the subjects about the experiments are available on the remote repository at:

https://github.com/jpiskozub?tab=repositories