Article:

Ayodele, E, Bao, T orcid.org/0000-0002-1103-2660, Zaidi, SAR orcid.org/0000-0003-1969-3727 et al. (4 more authors) (2021) Grasp Classification with Weft Knit Data Glove using a Convolutional Neural Network. IEEE Sensors Journal, 21 (9). pp. 10824-10833. ISSN 1530-437X

https://doi.org/10.1109/JSEN.2021.3059028

Grasp Classification with Weft Knit Data Glove using a Convolutional Neural Network

Emmanuel Ayodele[1], Tianzhe Bao[1], SAR Zaidi[1], Ali Hayajneh[2], Jane Scott[3], Zhiqiang Zhang[1], Des McLernon[1]

***Abstract*—Grasp classification using data gloves can enable therapists to monitor patients efficiently by providing concise information about the activities performed by these patients. Although, classical machine learning algorithms have been applied in grasp classification, they require manual feature extraction to achieve high accuracy. In contrast, convolutional neural networks (CNNs) have outperformed popular machine learning algorithms in several classification scenarios because of their ability to extract features automatically from raw data. However, they have not been implemented on grasp classification using a data glove. In this study, we apply a CNN in grasp classification using a piezoresistive textile data glove knitted from conductive yarn and an elastomeric yarn. The data glove was used to collect data from five participants who grasped thirty objects each following Schlesinger's taxonomy. We investigate a CNN's performance in two scenarios where the validation objects are known and unknown. Our results show that a simple CNN architecture outperformed k-nn, Gaussian SVM, and Decision Tree algorithms in both scenarios in terms of the classification accuracy.**

***Index Terms*—CNN, Data glove, Grasp classification, Knit strain sensors.**

I. Introduction

PROGRESS measurement is an important factor in the rehabilitation of patients. Conventionally, progress measurement is performed by a physiotherapist who manually checks the progress at the injured joint. This method is costly as it involves frequent travel by the patient or physiotherapist. Furthermore, the chance of a physiotherapist's visit coinciding with important progress events is very limited. Therefore, researchers have developed several approaches to solve this challenge. Particularly, all approaches can be categorised into two major methods. These methods are a) *Camera-based methods* and b) *Wearable devices*. Camera-based methods involve using cameras to detect motion at the joints of the patient and processing the data into relevant information [1]. Although, there have been successful applications of this approach in research studies, the commercial adoption of this method has been constricted by the fear of intrusion into the privacy of the patient [2]. In addition, the use of a camera-based method limits the movement of the patient to within the camera's view thus restricting the patient from performing their daily activities. In contrast, wearable devices can collect data from the affected joint without restricting the movement of the patient. Subsequently, the collected data is uploaded to a computer or the cloud where the physiotherapist can remotely monitor the progress of the patient. Moreover, this enables the physiotherapist to monitor the progress of multiple patients conveniently.

Wearable devices are worn by the user and therefore, face a weight constraint as they must be light weight to prevent further injuries to the affected joint. In the progress measurement of interphalangeal joints, the popular wearable device is a data glove. The conventional design of a data glove is to integrate a strain sensor into a textile data glove by a form of external attachment. This design method leads to bulky data gloves that are conspicuous and therefore, unappealing to patients. In addition, the degradation of this external attachment can cause inaccuracies in the glove's measurement.

The use of weft knit sensors in wearable devices provides a substantial potential in designing textile wearable devices that are light weight, flexible and accurate [3]. Wearable devices that comprise of weft knit sensors include a knee sleeve and a respiration belt [4], [5]. In our earlier work [6], we designed a lightweight textile data glove whose sensors and support structure are wholly textile. The entire glove is fabricated in a single manufacturing process thus eliminating the need for an external attachment between the support structure and the strain sensors. We achieved this by weft knitting conductive yarn and an elastomeric yarn into weft knit sensors and weft knitting the rest of the glove with the elastomeric yarn using

[1] Institute of Robotics, Autonomous Systems and Sensing; University of Leeds, Leeds LS2 9JT, UK. Syed Ali Raza Zaidi is the corresponding author (S.A.Zaidi@leeds.ac.uk).

[2] Department of Electrical Engineering. Faculty of Engineering. The Hashemite University. P.O. Box 330127, Zarqa 13133, Jordan.

[3] School of Architecture, Planning and Landscape, Newcastle University, Newcastle NE1 7RU, UK.

WholeGarment^{TM} technology. Consequently, our data glove provides the feel and appearance of normal clothing while being capable of sensing strain.

Classification of the acquired data into comprehensible information is vital for the increased adoption of wearable devices as it is impractical for physiotherapists to understand the raw data. The use of machine learning in conjunction with a data glove to classify acquired data into various sign languages is quite popular [7]–[9]. However, only a few studies have utilised machine learning techniques in classifying the grasps performed with a data glove. Particularly, Bernardin et al. [10] employed HMM to classify gestures made with a sensor fusion of tactile sensors and Cyberglove. The gestures were classified using Kamakura taxonomy into four major categories: power, intermediate, precision and thumbless grips. Classification accuracy was an average of 85.25% for the single-user system and 91.5% for the multiple-user system. In addition, Heumer et al. [11] compared 28 different classifiers categorised into Lazy, function approximators, Tree-based and Rules-based and Bayes classifiers in the classification of grasps performed using a Cyberglove. It was observed that on average, function approximating classifiers performed best with a minimum and maximum accuracy of 81.41% and 86.8% respectively. Although, the results of these classical machine learning algorithms are quite promising, they are limited by the selection of their hand-crafted features. The performances of these algorithms are limited because they rely on the manual selection of features that best represent the data.

In contrast, deep neural networks (DNN) extract optimal features directly from the data by its layer-by-layer processing and in-model feature transformation. This has enabled DNN to outperform classical machine learning techniques in various applications such as computer vision, speech recognition and disease detection [12]–[18]. Convolutional neural networks (CNNs) are the most popular DNN algorithms. Typically, they comprise of stacked convolutional filters, activation and pooling layers that enable its optimal selection of discriminative features in a time-series data. CNN algorithms have been very successful across several fields particularly in the field of rehabilitation using electrocardiography (ECG) and electromyography (EMG) data [19]–[22].

Furthermore, CNN algorithms have been employed in grasp classification, albeit using a camera-based method. Notably, images of 500 objects were classified into four categories: pinch, tripod, palmar wrist neutral and palmar wrist pronated. In an offline test, the CNN algorithm performed at an accuracy of 85% for seen objects and an accuracy of 75% for unseen objects [23]. Seen objects were objects used for the algorithm's validation that were included in the training data while unseen objects were validation objects that were not included in the training data and were therefore novel to the algorithm.

In addition, CNNs have been utilised successfully in other glove-based gesture classification. The taxonomies in these studies include sign languages and custom taxonomies [24]–[26]. In particular, CNN was used to classify hand poses acquired with a data glove [27]. The classification accuracy was computed to be 89.4%. However, the study was limited to only one participant.

Although CNN algorithms have performed excellently across several classification applications, to the best of our knowledge, they have not been implemented in grasp classification using a data glove. Therefore, in this paper, we propose applying CNN in classifying grasps performed with the weft knit data glove. We compare the results with popular classical machine learning algorithms. Our results show that the simple CNN architecture outperforms the classical machine learning algorithms. The structure of the rest of this paper is as follows. Section II describes the data acquisition hardware including the weft knit data glove and its sensor configuration. The CNN algorithm and the classification scenarios are reported in Section III. Sections IV, V and VI illustrate the results, discussion and conclusion respectively.

II. Data Acquisition

A. Weft Knit Sensor

The strain sensors are created by weft-knitting conductive yarn and an elastomeric yarn in a plain knit structure. Furthermore, we design a novel architecture (shown in Fig. 1) such that each course of loops from conductive yarn is accompanied by a course of loops from the elastomeric yarn. Particularly, the conductive yarn used is a multi-filament yarn comprising of 80% polyester and 20% stainless steel. It is a Schoeller multifilament conductive yarn commercially available from Uppingham Yarns Ltd. According to its specification sheet, it has a maximum extension of 5.5% and its resistivity varies between $(200 - 1800\Omega m)$ depending on the yarn tension. We selected a multifilament yarn instead of a coated yarn because coated yarns are subject to environmental degradation.

1) Electromechanical model: A simplified electromechanical model of the sensor is illustrated in Fig. 1 depicting the resistive circuit of a knit loop in the sensor. The circuit comprises of length resistances R_l and R_h that represent the resistance of the legs and heads/sinkers of the knit loop respectively. These resistances can be calculated as:

$$R_l = \frac{\rho L_l}{A_r}, \tag{1}$$

$$R_h = \frac{\rho L_h}{A_r}, \tag{2}$$

where A_r is the cross-sectional area of the conductive yarn. L_l and L_h are the lengths of the loop legs and loop head/sinker respectively as shown in Figure 1 and can be calculated using any of the several geometrical models of a knit loop [28]–[30].

The contact resistance is the major factor in the piezoresistivity of the weft knit sensor. According to Holm's contact theory, a contact resistance occurs when two conductors are in contact with each other. This contact resistance is dependent on the contact pressure between the conductors. The elasticity of the weft knit structure and the elastomer causes the contact pressure between the conductive yarn loops to change when it is extended. This contact pressure affects the contact resistance as shown in the Holm's contact resistance equation below:

$$R_c = \frac{\rho}{2}\sqrt{\frac{\pi H}{nP}}, \tag{3}$$

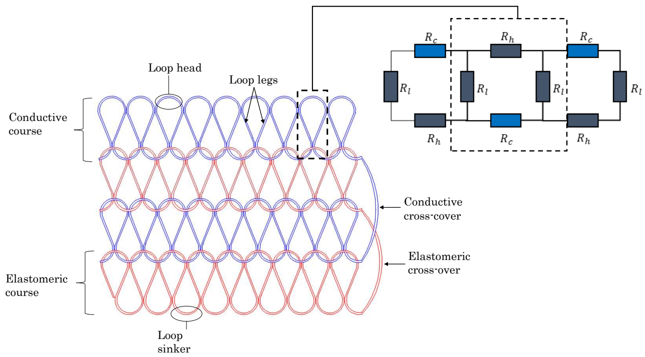

Fig. 1. Weft knit sensor design and its equivalent electrical circuit.

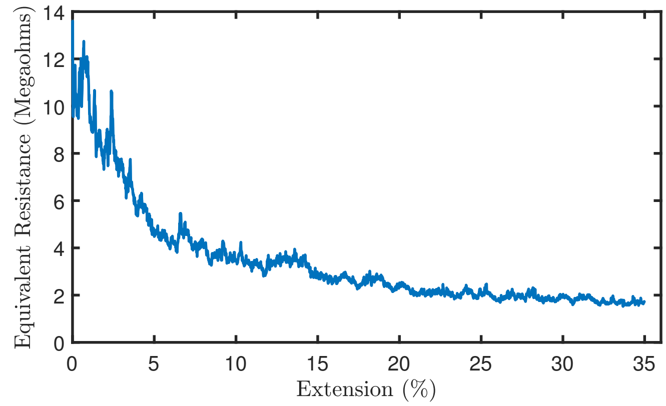

Fig. 2. Tensile test illustrating sensor's piezoresistivity.

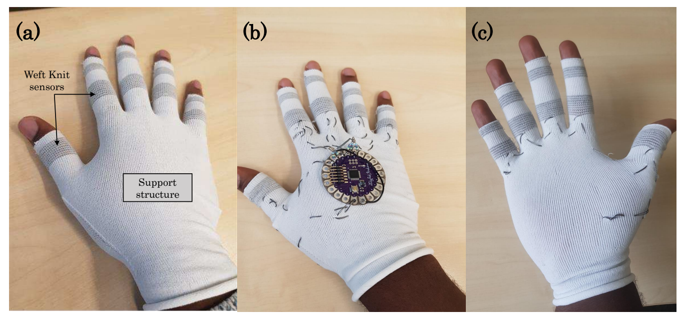

Fig. 3. (a) Fabricated weft knit data glove, (b) Front view of the data glove and its embedded measurement setup, and (c) Back view illustrating connection with conductive thread.

where, R_c is the contact resistance, ρ is the electrical resistivity, H is the hardness of the material used, n is the number of contact points and P is the contact pressure between the conductive loops. The equivalent (total) resistance of the sensor comprising of the contact resistances and the length resistances can be calculated using Kirchhoff's circuit analysis.

2) Sensor Characterisation: A strain test was performed to illustrate the electromechanical behaviour of the sensor configuration used in the glove. The experiment was performed using a tensile testing machine (Instron 3369) and a digital multimeter. Three sensors were knitted with 72 courses (row of knitted loops) and 36 wales (column of knitted loops). Due to the sensor's architecture, there were 36 courses of conductive yarn and 36 courses of elastomeric yarn. The sensors were stretched at a speed of 10mm/min until they reached 35% extension while their resistance was measured with a multimeter.

The average result of the tensile test is shown in Fig 2. It was observed that the sensor's resistance reduced exponentially as its extension increased. This occurred because as the sensor was extended, the contact pressure between the conducting loops increased thereby reducing the contact resistance and consequently, the equivalent resistance. The change in equivalent resistance reduced significantly as the extension of the sensor surpassed 25% because contact resistance between the loops was negligible due to the high contact pressure. This section is vital as it illustrates the electrical behaviour of the sensor as it is extended by movements at the interphalangeal joints. Furthermore, the results of the tensile test show that the sensor does not exhibit a perfectly linear piezoresistivity. The exponential piezoresistivity of the sensor may increase the difficulty in classifying acquired data.

B. Data Glove

The data glove illustrated in Fig. 3a is a wholly knitted textile glove with no external attachment between the support structure and the strain sensors. This was achieved by knitting the sensors and the support structure in a single fabrication process using WholeGarment^{TM} technology. Data is transmitted by sewing conductive thread from the sensors in the data glove to the analog-digital converters (ADC) located in the microprocessor (Arduino Lilypad). A voltage divider circuit enables the ADC to convert the resistance of the weft knit sensors to digital values between 0 and 1023. The microprocessor is connected to a computer (Intel I7-8750H, 16GB RAM, Nvidia GTX1060) for offline processing on MATLAB R2019. The USB port of the computer also powers the microprocessor. Furthermore, positive and negative connections are prevented from creating a short circuit by sewing the negative connections at the back of the glove and positive connections at the front of the glove. The measurement setup is depicted in Fig. 3(b) and (c).

C. Experimental Setup

This study was approved by the Faculty Research Ethics Committee of University of Leeds, UK (reference: MEEC 19-006). There were five healthy participants in this study including three males and two females. All participants signed an informed consent form.

The Schlesinger taxonomy [31], [32] was used in this study for selecting the grasp types. This taxonomy is widely known to be the earliest study to accurately categorise the different grasps of a human hand [33]. We selected this taxonomy as a research constraint that acts as a base in which more patient-tailored taxonomies can be built upon.

For each grasp type shown in Fig. 4, 5 objects were selected for the experiment. These objects and their corresponding grasp type are enumerated in Table I. The participants performed five grasps per object thereby providing a total of 750 samples (5 participants x 5 grasps x 30 objects). Each

TABLE I
OBJECTS USED IN THE EXPERIMENT AND THEIR GRASP TYPES

Grasp Type	Objects
Cylindrical	Water bottle, flask, coffee cup, can, plastic bottle.
Hook	Mug, bag strap, headphones, kettle, back pack.
Lateral	Key, CD, ruler, id card, spoon.
Palmar	breadboard, phone, match box, multimeter, plastic case.
Spherical	Lemon, orange, apple, mouse, onion.
Tip	Pen, pencil, chopstick, stylus, ball pen.

grasp was for 30 seconds and participants were allowed to take breaks during the experiment to prevent fatigue.

III. Deep Learning Approach

A. Data Pre-processing

Data was recorded by the glove at a frequency of 20 hertz from the five sensors located at the distal interphalangeal joints. For each 30 seconds grasp of an object, 3000 (600 x 5 sensors) data values were recorded. This data obtained in the time series represents the signal features. As CNN requires a 3d image as an input, each grasp is represented as a 600x5x1 array. In this array, the first dimension (600 elements) represents the acquisition of 30 seconds of data at 20 hertz from each sensor while the second dimension (5 elements) represents the five sensors that transmit data to the microprocessor. Furthermore, the temporal order in which the data was acquired was unaltered. A short transition time was implemented between each new grasp to facilitate the collection of data. This transition time was later removed from the data to ensure that only the grasping period was recorded from the glove. In addition, this eliminated the complexities that involve the starting position of the grasping hand.

We perform no feature extraction or filtering of the data for CNN or the classical machine learning algorithms as this study aims to show the performance of algorithms in classifying raw data from weft knit sensors. Particularly, as research on classification using weft knit sensors is still nascent, it would be impractical to extract features manually.

B. CNN Algorithm

Convolutional Neural Networks are feed forward deep neural networks consisting of stacks of convolutional and pooling layers and then one or more fully connected layers [34], [35]. The convolutional layers employ convolution in extracting the features from the input data. Particularly, feature maps are generated by convolving the input signal with filters (kernels) consisting of neurons with learnable weights and biases. The convolution operation of the g-th feature map on the f-th convolutional layer located at position (a, b) can be described as:

$$v_{f,g}^{a,b} = \sigma\left(b_{f,g} + \sum_{i}\sum_{x=0}^{X_f-1}\sum_{y=0}^{Y_f-1} w_{f,g,i}^{x,y} v_{f-1,i}^{a+x,b+y}\right), \quad (4)$$

where $b_{f,g}$ is the feature map's bias, $w_{f,g,i}^{x,y}$ is from the weight matrix, X and Y are the kernel's height and width respectively, and $\sigma(\cdot)$ is a non-linear activation function such as Rectified Linear Unit (RELU), Sigmod or Tanh. In our architecture we use a RELU non-linear function and it can be represented as:

$$\sigma(k) = \max(0, k). \quad (5)$$

A pooling layer is added between convolutional layers to increase the invariance of the feature maps to minor changes in the input. It achieves this by aggregating the neighbouring outputs as a representative of the spatial region. In earlier studies, average pooling was the standard. However, maximum pooling has become the benchmark in state-of-the-art CNN approaches [34]. Similar to traditional neural networks, the fully-connected (FC) layer(s) classifies the input signal based on the extracted features obtained from previous layers.

C. CNN Architecture

An ablation study was performed to determine the optimal CNN configuration. Four parameters (i.e. the number of convolutional blocks, the number and size of convolutional filters, and the dropout layer's probability) were varied to create 16 CNN configurations. These parameters are known to significantly impact the performance of a CNN [36]. The configurations and their parameters are shown in Table II. All other parameters were constant for all configurations. In particular, each convolutional block had a rectified unit layer (RELU) acting as a nonlinear activation function, a downsampling pooling layer with filters of size 2x1 and a dropout layer to reduce overfitting. The last convolutional block was connected to a fully-connected layer with 6 hidden units representing the 6 grasp types, a softmax layer which employs a cross entropy loss function and a classification layer. Moreover, the networks were trained at a dynamic learning rate using stochastic gradient descent. The initial learning rates were 0.001 and were reduced by 95% after every 10 epochs. The batch sizes were fixed at 16 and the number of epochs was 36.

These configurations were utilised in classifying the data in two experiments. In the first experiment, one grasp was used as the validation data while the remaining 4 grasps were used as the training data i.e (80% training data and 20% validation data). Thereafter, cross validation was performed by repeating the experiment 5 times where each grasp was utilised as the validation data. In the second experiment, the CNN configurations were trained with 4 out of 5 objects with the remaining object as the validation data i.e (80% training data and 20% validation data). Cross validation was also performed by repeating the experiment 5 times where each object was used as the validation data. The average accuracy of each CNN classifier in both experiments was calculated. These experiments were performed on Participant 1's data with the aim of utilising the best CNN configuration in terms of classification accuracy on an expanded experiment comprising of all participants.

The results of this study are also shown in Table II. It was observed that CNN configurations with two convolution blocks had a higher accuracy than similar configurations with

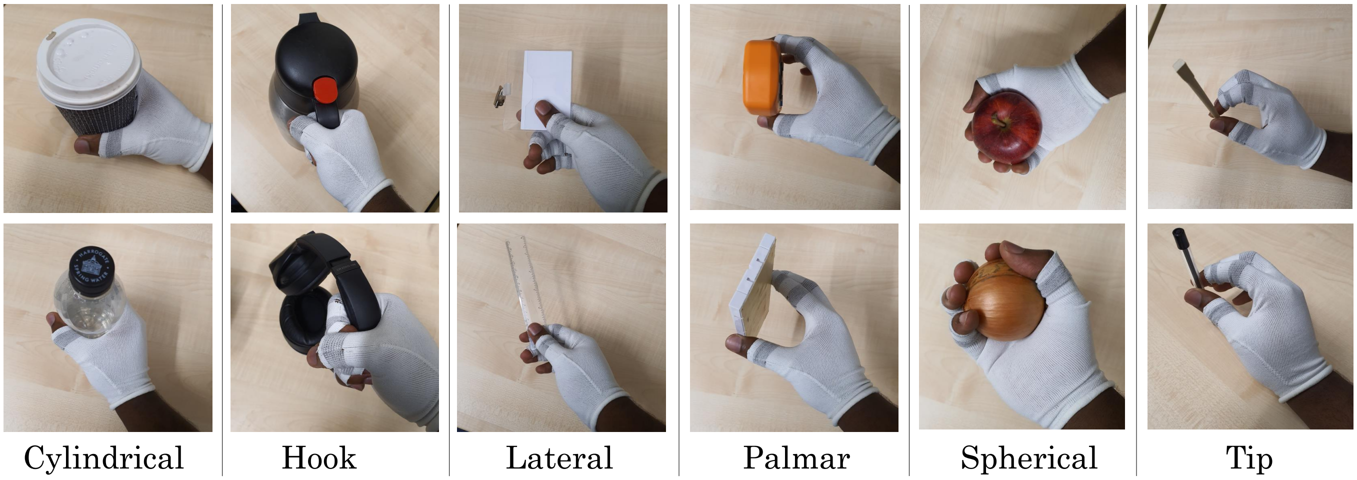

Fig. 4. Grasp types of objects used in the study (Schlesinger taxonomy).

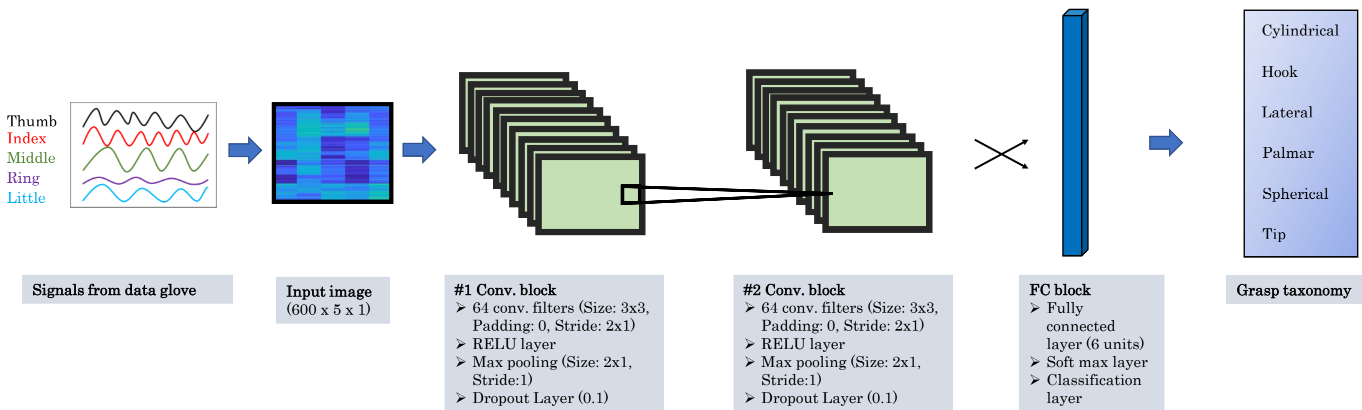

Fig. 5. CNN architecture (C15) for grasp classification.

TABLE II
CNN CONFIGURATIONS AND THEIR RESPECTIVE PARAMETERS.

Config.	Conv. Filter size	No of Conv. Filters	No of Conv. blocks	Dropout Probability	Accuracy	Run time
C1	3x2	32	1	0.1	81.00	4.5s
C2	3x2	32	1	0.2	82.67	4.4s
C3	3x2	32	2	0.1	83.67	6.0s
C4	3x2	32	2	0.2	82.00	6.1s
C5	3x2	64	1	0.1	79.67	5.0s
C6	3x2	64	1	0.2	76.33	5.0s
C7	3x2	64	2	0.1	83.67	6.7s
C8	3x2	64	2	0.2	81.00	6.9s
C9	3x3	32	1	0.1	78.67	4.7s
C10	3x3	32	1	0.2	80.67	5.0s
C11	3x3	32	2	0.1	83.67	6.3s
C12	3x3	32	2	0.2	82.00	6.1s
C13	3x3	64	1	0.1	81.00	5.1s
C14	3x3	64	1	0.2	78.33	5.1s
C15	3x3	64	2	0.1	**86.00**	6.2s
C16	3x3	64	2	0.2	82.34	6.3s

only one convolutional block. However, the higher accuracy occurred at a computation cost as observed in the increased run times seen in configurations with two convolutional blocks. In particular, configurations with two convolutional blocks had run times that were on average 1.5 seconds longer than similar configurations. However, the aim of this ablation study was to select the optimal CNN configuration in terms of its accuracy. Therefore, classifier C15 illustrated in Fig. 5 was seen to achieve the highest average classification accuracy and was selected as the optimal CNN configuration. Moreover, in comparison with configurations with two convolutional blocks, the computation time of C15 was relatively low. No further optimisation of C15 was performed in its implementation on the expanded experiment. This study was important in ensuring that the optimal parameters were selected for the CNN algorithm.

D. Classification Scenarios

In this study, we evaluate the performance of the selected CNN (C15) and other algorithms on the following classification scenarios. These scenarios are:

1) Object seen: This scenario exemplifies applications where the validation objects are known. That is, the objects in the validation data are part of the training data. Traditionally, classifiers will achieve high accuracy in this scenario but because weft knit sensors experience hysteresis and drift, the performance of the classifiers will be adversely affected. In this scenario, the classifiers were trained with 4 out of 5 grasps of an object and validated with the last grasp of the object (i.e. 120 images for training and 30 images for validation per participant). Cross validation was performed by repeating this experiment 5 times where each grasp of an object was selected as the validation data and computing the average accuracy. Furthermore, this was repeated for all participants and the average accuracy was recorded.

2) Object unseen: This scenario illustrates applications where the objects grasped by the patient are unknown. It ensures that the therapist is provided with some information about the grasp type despite the object being held by the patient is not part of the training data set. In these experiments, the classifiers were trained with 4 out of the 5 objects in each grasp type and were validated with the last object (120 images for training and 30 images for validation per participant). Similar to the object seen experiment, cross validation was performed by repeating the experiment 5 times where each object was selected as the validation data and the average accuracy was computed. In addition, the experiment was repeated for all participants.

E. Comparative Machine Learning Techniques

In this study, popular machine learning techniques were implemented to compare their performance with the CNN in the various applications. These techniques include k-nearest neighbours (k-nn), Support Vector machine (SVM) and Decision Trees (trees) [37]–[41]. The default parameters in Matlab R2019's Machine Learning Toolbox were selected for the various configurations of these techniques. As there are no classification studies with weft knit sensors, these parameters were chosen from a popular and reliable toolbox to provide a verifiable comparative study.

1) k-nearest neighbours (k-nn): k-nn is a probabilistic pattern recognition technique that classifies a signal output based on the most common class of its k nearest neighbours in the training data. The most common class (also referred to as the similarity function) can be computed as a distance or correlation metric. In this study, we select the Euclidean distance as the similarity function as it is the most commonly used metric in k-nn. The number of k-neighbours was varied to be 1, 10 and 100 for fine, medium and coarse k-nn techniques respectively. The probability density function $p(\mathbf{M}, c_j)$ of the output data $\mathbf{M}$ belonging to a class c_j with jth training categories can be computed as:

$$p(\mathbf{M}, c_j) = \sum_{n_z \epsilon knn} d(\mathbf{M}, n_z) V(n_z, c_j), \tag{6}$$

where n_z is a neighbour in the training set, $V(n_z, c_j)$. The Euclidean distance $d(\mathbf{M}, n_z)$ of output data $\mathbf{M}$ and neighbour n_z can be calculated as:

$$d(\mathbf{M}, n_z) = \sqrt{\sum_{z=1}^{k} (\mathbf{M}_z - n_z)^2}. \tag{7}$$

2) Gaussian SVM: Traditionally, support vector machines (SVM) is a supervised learning method used for performing linear classification. However, the data obtained during experiment cannot be separated using linear hyperplanes because of the close resemblance of some grasp types and the hysteresis and drift that occur in a weft knit strain sensor. In order to use SVMs for non-linear classification, we apply Gaussian kernels which can map the data into an unlimited dimension space. Three variations of Gaussian SVM were implemented by selecting 7.9, 32, and 130 on the kernel scale for fine, medium and coarse Gaussian SVM respectively. The decision function for Gaussian SVM classification of pattern data $\mathbf{u}$ can be represented as:

$$f(\mathbf{u}) = \text{sign}\left(\sum_{k=1}^{h} \lambda_k c_k \exp\left(\frac{-\|\mathbf{u}_k - \mathbf{u}\|^2}{2\sigma^2}\right) + t \right), \tag{8}$$

where c_k is the class label for the k-th support vector $\mathbf{u}_k$, λ_k is the Lagrange multiplier, and t is the bias.

3) Decision Tree: Decision tree is a supervised learning technique that aims to split classification into a set of decisions that determine the class of the signal. The output of the algorithm is a tree whose decision nodes have multiple branches and its leaf nodes deciding the classes. Three configurations of the Decision tree algorithm were implemented by varying the maximum number of splits as 100, 20 and 4 for fine, medium and coarse Decision tree respectively.

TABLE III
ACCURACY OF CNN CLASSIFIER FOR EACH PARTICIPANT IN THE TWO CLASSIFICATION SCENARIOS

Participants	Object seen		Object unseen	
	Mean	**Std.**	**Mean**	**Std.**
P1	91.33	2.66	76.00	4.90
P2	87.33	9.29	74.00	13.40
P3	80.67	4.90	69.33	12.54
P4	82.67	6.80	66.67	8.69
P5	99.33	1.33	92.67	9.98
Average	**88.27**	**5.00**	**75.73**	**9.90**

TABLE IV
ACCURACY OF THE CLASSIFIERS IN THE TWO CLASSIFICATION SCENARIOS. THE BEST CLASSIFIER IS HIGHLIGHTED WITH A BOLD FONT.

Classifier	Object seen		Object unseen		Run time
	Mean	**Std.**	**Mean**	**Std.**	
Fine k-nn	83.87	10.30	69.47	14.63	0.86s
Medium k-nn	77.07	8.65	69.07	8.93	0.85s
Coarse k-nn	32.53	7.53	30.80	6.72	0.85s
Fine SVM	39.60	6.79	27.07	5.88	1.39s
Medium SVM	82.80	8.13	70.53	10.52	1.34s
Coarse SVM	79.20	8.81	70.27	11.82	1.32s
Fine tree	68.13	10.06	58.40	12.42	0.92s
Medium tree	68.13	10.06	58.40	12.42	0.95s
Coarse tree	57.47	7.72	53.47	8.24	0.90s
CNN	**88.27**	5.00	**75.73**	9.90	6.20s

IV. RESULTS

A. Object seen

Fig. 7 illustrates the accuracy of the classifiers when the object to be grasped is known. CNN outperforms all the classical classifiers with an average accuracy of 88.27%. This accuracy is slightly lower than results obtained by commercial data gloves in other classification scenarios. This is caused by the drift that occurs in weft knit sensors. Drift causes the

Fig. 6. Confusion matrix depicting the average results of the object seen scenario.

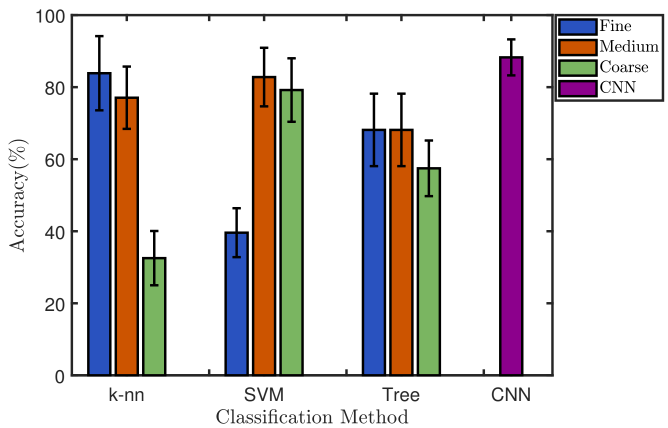

Fig. 7. Object seen. Bars represent mean accuracy of the classifier and error-bars illustrate the standard deviation.

output of the sensor to stray despite the absence of change in its extension.

Fig. 6 illustrates the confusion matrix of the average results of all participants in the object seen scenario. The confusion matrix shows that grasps of Hook, Lateral, Spherical and Tip are classified excellently at 87.2%, 92.8%, 92% and 92.8% respectively. In contrast, the average classification accuracy of Cylindirical and Palmar grasps were significantly lower at 80% and 84.8% respectively.

Fig. 8 depicts a detailed view of the average classifier class performance on each participant. CNN outperforms all classifier classes for each participant in terms of its mean accuracy. In particular, it outperforms other classifier classes by an average of 21% in terms of its mean classification accuracy.

B. Object unseen

Fig. 10 depicts the accuracy of the classifiers when the validation object is unknown. This exemplifies applications where the glove may be used to grasp objects not within the training data. It was observed that the accuracy of the classifiers in this scenario were lower than the accuracy seen in object seen scenario. This was expected as it is common in glove-based gesture classification because the validation objects are not

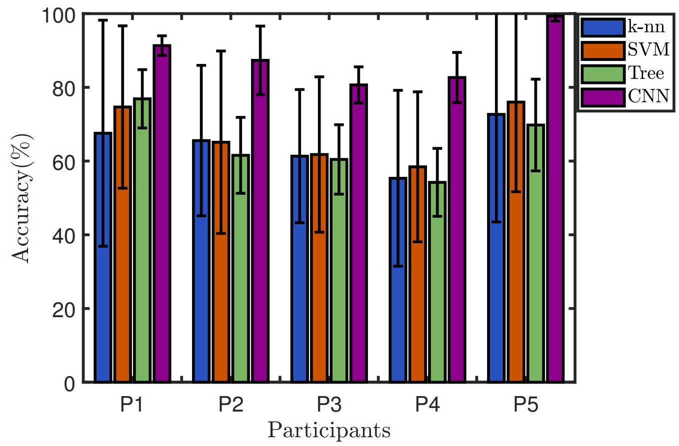

Fig. 8. Detailed results of object seen. Bars represent mean accuracy of the classifier class performance for each participant and error-bars illustrate the standard deviation.

Fig. 9. Confusion matrix depicting the average results of the object unseen scenario.

part of the training data (i.e., they are unknown). Nonetheless, CNN outperforms the classical machine learning methods with an average accuracy of 75.73%.

Fig. 11 illustrates an expanded view of the performance of each classifier class on the participants. CNN outperforms other classifier classes in each participant in terms of its mean accuracy. Particularly, for P5, it outperforms the next best classifier class by 23.8%.

Fig. 9 depicts the confusion matrix of the average results of all participants in the object unseen scenario. Similar to the results obtained in the object seen scenario, the algorithm struggled with classifying Cylindrical and Palmar objects with classification accuracy of 63.2% and 68.8% respectively. In contrast, higher classification accuracy were achieved in Hook, Lateral, Spherical and Tip objects with accuracy of 72.8%, 84%, 79.2% and 86.4% respectively.

V. Discussion

In the last decade, the implementation of convolutional neural networks in several applications has been very popular. These applications include image and text classification, disease recognition and gait classification. In these applications, CNN has outperformed popular machine learning algorithms

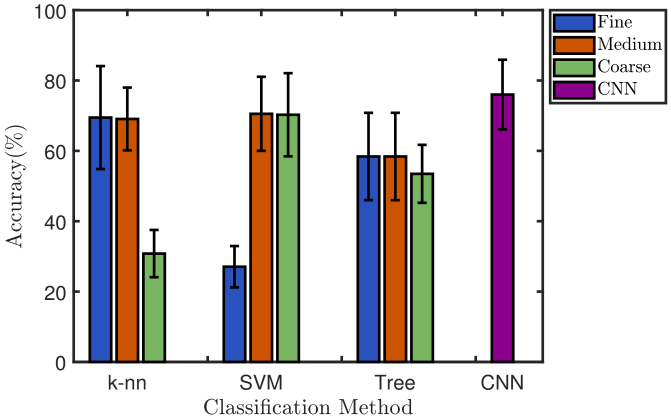

Fig. 10. Object unseen. Bars represent mean accuracy of the classifier and error-bars illustrate the standard deviation.

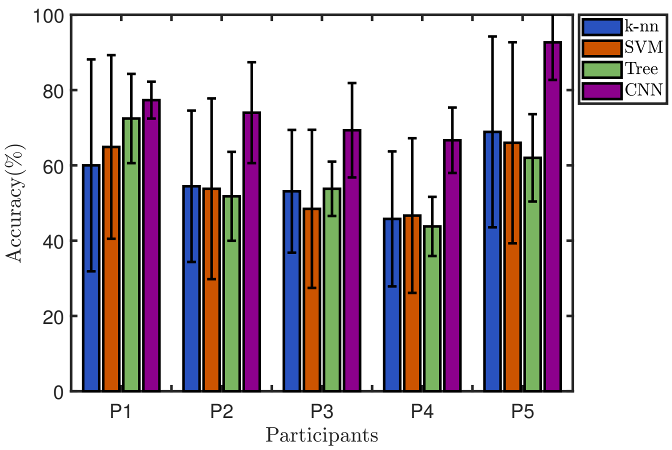

Fig. 11. Detailed results of object unseen. Bars represent mean accuracy of the classifier class performance for each participant and error-bars illustrate the standard deviation.

because of its ability to automatically extract features from the data set. In contrast, machine learning algorithms require manual feature extraction techniques such principal component analysis or dimensionality reduction to produce accurate classification accuracy. However, despite its popularity, there has been no research on its application to grasp classification from data obtained with a piezoresistive data glove. Therefore, this study aims to bridge that gap by implementing a CNN architecture that outperforms classical machine learning algorithms in this application.

Our results show that a simple CNN architecture outperforms k-nn, Gaussian SVM and Decision Tree algorithms in both classification scenarios. Moreover, the simplicity of our CNN architecture is intentional. Particularly, the absence of research illustrating the implementation of CNNs in this application caused us to investigate the performance of a simple architecture before applying more complex CNN architectures. However, the computation cost of CNN was higher than the comparative algorithms as seen in the run times shown in Table IV. This was expected as CNN and deep learning algorithms are known for their higher computational costs as a result of their automatic feature extraction.

In addition, the results in Table III illustrate that the accuracy of all algorithms are higher for P5 (participant 5) than for other participants. This transpired because the data glove was created to fit the hand size of this participant. This illustrates the potential of textile wearables, as the one-size-fits-all constraints can be eliminated by fabricating these devices alongside the conventional size measurements (for example: XS-extra small, S-small, M-medium, L-large etc.) that have been used in the clothing industry for several decades. Therefore, by utilising weft knit sensors, higher classification accuracy can be achieved by creating perfectly fitting wearables based on the user's physical dimensions.

Furthermore, the results of this study in Table IV show that the average accuracy of most classifiers reduced in the second classification scenario. This scenario depicted an application of the glove where the grasp type of the object is unknown. Consequently, the validation data set comprises objects not in the training data set. Therefore, it is a more difficult classification problem for the algorithms. However, despite this difficulty, CNN still outperforms other classifiers.

Although, CNN outperforms other classifiers, its average accuracy among the participants is less than 90%. However, we have shown that for participants for whom the glove is specifically designed for, then the average accuracy was much higher (>99% for seen objects and >92% for unseen objects) regardless of whether the validation object was part of the training set. This is remarkable for classification using weft knit sensors as they are still technologically immature and struggle with hysteresis and drift. This is a fertile area for further research as more deep learning architectures such as LSTM (long short-term memory) or CNN-LSTM can be applied in the classification of their raw data. Recently, a study illustrated the use of LSTM on grasp classification using a knitted glove [42]. It will be interesting to compare the performance of CNN to LSTM in grasp classification from data acquired with a knitted data glove. Although the memory properties of LSTM should provide an advantage over CNN [26], CNN has also been seen to outperform LSTM [24]. Therefore, it will be interesting to see if more complex deep learning algorithms improve the accuracy of grasp classification using data gloves. Higher performances (>95% average accuracy) in this application may rapidly increase the commercial adoption of data gloves in rehabilitation.

VI. Conclusion

In this paper, we have pioneered the use of convolutional neural networks on grasp classification using a piezoresistive data glove. Our simple CNN architecture consisting of only two convolutional blocks outperformed classical machine learning techniques in the two classification scenarios. Notably, the average classification accuracy of our CNN algorithm was 88.27% and 75.73% in the object seen and object unseen scenarios respectively. Future work will involve the application of more robust deep learning approaches such as RNN and CNN-LSTM to improve the accuracy in gesture prediction applications using a larger dataset of participants.

References

[1] R. Y. Wang and J. Popović, "Real-time hand-tracking with a color glove," *ACM transactions on graphics (TOG)*, vol. 28, no. 3, p. 63, 2009.

[2] K. E. Caine, A. D. Fisk, and W. A. Rogers, "Benefits and privacy concerns of a home equipped with a visual sensing system: A perspective from older adults," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 2. SAGE Publications Sage CA: Los Angeles, CA, 2006, pp. 180–184.

[3] H. Zhang, X. Tao, S. Wang, and T. Yu, "Electro-mechanical properties of knitted fabric made from conductive multi-filament yarn under unidirectional extension," *Textile research journal*, vol. 75, no. 8, pp. 598–606, 2005.

[4] Y. Li, X. Miao, and R. K. Raji, "Flexible knitted sensing device for identifying knee joint motion patterns," *Smart Materials and Structures*, vol. 28, no. 11, p. 115042, 2019.

[5] O. Atalay, W. R. Kennon, and E. Demirok, "Weft-knitted strain sensor for monitoring respiratory rate and its electro-mechanical modeling," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 110–122, 2015.

[6] E. Ayodele, S. Zaidi, Z. Zhang, J. Scott, Q. Kong, and D. Mclernon, "Weft knit smart data glove," in *IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2019.

[7] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous arabic sign language recognition in user-dependent mode," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 526–533, 2015.

[8] J.-S. Kim, W. Jang, and Z. Bien, "A dynamic gesture recognition system for the korean sign language (ksl)," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 26, no. 2, pp. 354–359, 1996.

[9] S. A. Mehdi and Y. N. Khan, "Sign language recognition using sensor gloves," in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02.*, vol. 5. IEEE, 2002, pp. 2204–2206.

[10] K. Bernardin, K. Ogawara, K. Ikeuchi, and R. Dillmann, "A sensor fusion approach for recognizing continuous human grasping sequences using hidden markov models," *IEEE Transactions on Robotics*, vol. 21, no. 1, pp. 47–57, 2005.

[11] G. Heumer, H. B. Amor, M. Weber, and B. Jung, "Grasp recognition with uncalibrated data gloves-a comparison of classification methods," in *2007 IEEE Virtual Reality Conference*. IEEE, 2007, pp. 19–26.

[12] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[14] F. Seide, G. Li, and D. Yu, "Conversational speech transcription using context-dependent deep neural networks," in *Twelfth annual conference of the international speech communication association*, 2011.

[15] G. Dahl, A.-r. Mohamed, G. E. Hinton *et al.*, "Phone recognition with the mean-covariance restricted boltzmann machine," in *Advances in neural information processing systems*, 2010, pp. 469–477.

[16] L. Deng, M. L. Seltzer, D. Yu, A. Acero, A.-r. Mohamed, and G. Hinton, "Binary coding of speech spectrograms using a deep auto-encoder," in *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[17] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Frontiers in neurorobotics*, vol. 10, p. 9, 2016.

[18] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[19] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, "Multi-class arrhythmia detection from 12-lead varied-length ecg using attention-based time-incremental convolutional neural network," *Information Fusion*, vol. 53, pp. 174–182, 2020.

[20] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ecg recordings," *IEEE journal of biomedical and health informatics*, vol. 22, no. 6, pp. 1744–1753, 2018.

[21] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2015.

[22] Z. Qin, Z. Jiang, J. Chen, C. Hu, and Y. Ma, "semg-based tremor severity evaluation for parkinson's disease using a light-weight cnn," *IEEE Signal Processing Letters*, vol. 26, no. 4, pp. 637–641, 2019.

[23] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "Deep learning-based artificial vision for grasp classification in myoelectric hands," *Journal of neural engineering*, vol. 14, no. 3, p. 036025, 2017.

[24] B. Fang, Q. Lv, J. Shan, F. Sun, H. Liu, D. Guo, and Y. Zhao, "Dynamic gesture recognition using inertial sensors-based data gloves," in *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 2019, pp. 390–395.

[25] N. Diliberti, C. Peng, C. Kaufman, Y. Dong, and J. T. Hansberger, "Real-time gesture recognition using 3d sensory data and a light convolutional neural network," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 401–410.

[26] M. A. Simão, O. Gibaru, and P. Neto, "Online recognition of incomplete gesture data to interface collaborative robots," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9372–9382, 2019.

[27] S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik, "Learning the signatures of the human grasp using a scalable tactile glove," *Nature*, vol. 569, no. 7758, pp. 698–702, 2019.

[28] F. T. Peirce, "The geometry of cloth structure," *Journal of the Textile Institute Transactions*, vol. 28, no. 3, pp. T45–T96, 1937.

[29] S. Kawabata, "Nonlinear mechanics of woven and knitted materials," in *Textile structural composites*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam: Elsevier, 1989, ch. 3.

[30] R. Postle and D. Munden, "Analysis of the dry-relaxed knitted-loop configuration: Part i: Two-dimensional analysis," *Journal of the Textile Institute*, vol. 58, no. 8, pp. 329–351, 1967.

[31] R. J. Schwarz and C. Taylor, "The anatomy and mechanics of the human hand," *Artificial limbs*, vol. 2, no. 2, pp. 22–35, 1955.

[32] G. Schlesinger, "Ersatzglieder und arbeitshilfen fur kriegsbeschadigte und unfallverletzte," *Der Mechanische Aufbau der Künstlichen Glieder*, pp. 312–661, 1919.

[33] J. Liu, F. Feng, Y. C. Nakamura, and N. S. Pollard, "A taxonomy of everyday grasps in action," in *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2014, pp. 573–580.

[34] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[35] Y. Zhao and S. Zhou, "Wearable device-based gait recognition using angle embedded gait dynamic images and a convolutional neural network," *Sensors*, vol. 17, no. 3, p. 478, 2017.

[36] B.-Y. Su, J. Wang, S.-Q. Liu, M. Sheng, J. Jiang, and K. Xiang, "A cnn-based method for intent recognition using inertial measurement units and intelligent lower limb prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1032–1042, 2019.

[37] D. Lu, Y. Yu, and H. Liu, "Gesture recognition using data glove: An extreme learning machine method," in *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2016, pp. 1349–1354.

[38] A. Ibarguren, I. Maurtua, and B. Sierra, "Layered architecture for real time sign recognition: Hand gesture and movement," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 7, pp. 1216–1228, 2010.

[39] N. S. Altman, "An introduction to kernel and nearest-neighbor non-parametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[40] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[41] X. Yang, X. Chen, X. Cao, S. Wei, and X. Zhang, "Chinese sign language recognition based on an optimized tree-structure framework," *IEEE journal of biomedical and health informatics*, vol. 21, no. 4, pp. 994–1004, 2016.

[42] J. Hughes, A. Spielberg, M. Chounlakone, G. Chang, W. Matusik, and D. Rus, "A simple, inexpensive, wearable glove with hybrid resistive-pressure sensors for computational sensing, proprioception, and task identification," *Advanced Intelligent Systems*, 2020.