European Public & Social Innovation Review

ISSN 2529-9824

Research article

Machine learning for grasping recognition using wearable sensors

Aprendizaje automático para el reconocimiento del tipo de agarre mediante sensores portables

Graciela Rodríguez Vega[1]: Universidad de Sonora, México.
gracirela.rodriguez@unison.mx
Dora A. Rodríguez Vega: Universidad Politécnica de Sinaloa, México.
drodriguez@upsin.edu.mx

Receipt Date: 05/06/2024

Acceptance Date: 15/10/2024

Publication Date: 30/01/2025

How to cite the article

Rodríguez Vega, G., & Rodríguez Vega, D. A. (2025) Machine learning for grasping recognition using wearable sensors [Aprendizaje de máquina para el reconocimiento del tipo de agarre mediante sensores portables]. *European Public & Social Innovation Review*, 10, 01-18. https://doi.org/10.31637/epsir-2025-1205

Abstract
Introduction: The objective is to evaluate the traditional classifiers for the identification of the grasp while doing different jobs, in order to obtain information that can be used in the diagnostic of the physical work requirements and job design. **Methodology:** The analysis considered different combinations of the data acquired from inertial and force resistive sensors: a) acceleration and resistive force sensors, b) acceleration, angular velocity and resistive force sensors c) acceleration, angular velocity, magnetic fields, and resistive force sensors. Different combinations of window and step sizes were selected with two overlap options: 50% and greater than 50%. Traditional classification models were trained: support vector machines, ensembles, Naive-Bayes algorithm. **Results:** Results demonstrate that the window size that presented optimal performance in the present study was 3 seconds with an overlap greater than 50%, the window size is greater than that suggested in the literature, which ranges from 0.75 to 2.25 seconds. **Conclusions:** The accuracy and F-score metrics for the different window-step combinations are presented, both metrics indicate that the models trained through Support Vector Machine have the best performance (90 %) with the combination of acceleration, angular velocity, and resistive force sensor.

[1] **Corresponding author:** Graciela Rodríguez Vega. Universidad de Sonora (México).

Keywords: Grasp recognition; hand motion analysis; wearable sensors; windowing; human behavior inference; segmentation; traditional classifiers; F-Score.

Resumen: Introducción: El objetivo del presente es evaluar el desempeño de clasificadores tradicionales en la identificación del tipo de agarre, para obtener información de interés en el diagnóstico de los requerimientos de trabajo físico y el diseño del trabajo. **Metodología**: El análisis considera diferentes combinaciones de datos adquiridos mediante sensores inerciales y resistivos de fuerza: a) aceleración y sensores resistivos de fuerza, b) aceleración, velocidad angular y sensores resistivos de fuerza, y c) aceleración, velocidad angular, campos magnéticos y sensores resistivos de fuerza. Diferentes combinaciones de tamaño de ventana y paso fueron seleccionados con dos opciones de traslape: 50% y mayor al 50%. Se entrenaron modelos de clasificaciones tradicionales: máquinas de vectores de soporte, ensambles y algoritmo Naive-Bayes. **Resultados:** Los resultados demostraron que el tamaño de ventana para un desempeño óptimo en el presente estudio es 3 segundos, con traslape mayor al 50%. Este tamaño de ventana es mayor al sugerido por la literatura (0.75 segundos-2.25 segundos)**. Conclusiones**: Se presentan la exactitud y el valor-F para las diferentes combinaciones de ventana-paso, indicando que los modelos estrenados mediante Máquinas de Vectores de Soporte presentaron el mejor desempeño (90%) al analizar la aceleración, la velocidad angular y los sensores resistivos de fuerza.

Palabras clave: Reconocimiento del agarre; análisis del movimiento de la mano; sensores portables; ventanas; inferencia del comportamiento humano; segmentación; clasificadores tradicionales; valor -F.

1. Introduction

The manual handling of objects is a common task in industrial activities that represents an ergonomic risk factor for the occurrence of musculoskeletal disorders. The upper extremities are frequently used during the performance of the daily activities of the human being, as well as in the work activities of the industrial sector, industrial activities are characterized by being highly repetitive, implying inappropriate postures, in addition to demanding the application of high forces, sometimes exceeding the physical capacities of the workers (Armstrong et al., 1982; Nur et al., 2014); this can cause various musculoskeletal disorders (MSDs), which can cause a decrease in work rate and generate productivity losses of about 7% in the United States (Nur et al., 2014) and about 2% of the gross domestic product of the nation. European Union (Bevan, 2015). In 2001, the Bureau of Labor Statistics of the United States reported that more than half of the companies that reported the occurrence of MSDs due to work, are prone to the development of these due to the use of hand tools and the intense hand work of the workers. It was also documented that carpal tunnel syndrome (one of the most common work-related MSDs of the hand) and general hand injuries generated an average of 25 and 13 days of absenteeism, respectively (Barr et al., 2004), strongly impacting the productivity of companies. Nur et al. (2014) concluded that 49.3% of MSD symptoms occur in the neck, 48% in the hand-wrist region, 46.7% in the shoulder, 33.6% in the back, and 21.7% in the lower back. According to studies in the food, transport, shipyards and textile industries of the European Union, published between 2000 and 2010, MSDs associated with the hand-wrist region have a prevalence of 42% (Govaerts et al., 2021) . The foregoing reveals the need to improve working conditions that generate health risks, mainly in the neck and hand-wrist area.

The objective of this study is to evaluate the performance of some traditional classification algorithms for the grasp type recognition, in addition to analyzing different options for selecting variables and segmenting the data flow from inertial sensors with 9 degrees of freedom (DOF), and resistive force sensors. The grasp type recognition, used in work activities is of great importance for the task design, thereby allowing the implementation of efficient work methods that reduce the adoption of unsuitable postures at work, which reduces the risk of musculoskeletal injury.

1.1. Grasp Type Classifications

The hand is one of the most complex motor systems in the human body (Xue et al., 2019). The study of the way in which humans take control of objects is important in the diagnosis of ergonomic risk factors in workplaces. The grasp type is defined as each static posture that the hand adopts to grasp a stable object during manipulation activities with one hand, regardless of its orientation (Barr et al., 2004; Feix et al., 2016).

Since the control of the object is an activity selected consciously or unconsciously by the worker, the recognition of the manipulation of objects becomes complicated (Xue et al., 2019). Feix et al. (2016) developed the GRASP Taxonomy of human grasp types, the proposed grasp categories including power, intermediate and precision grasp; in the power grasp all movements have to be evoked by the arm, while in the precision grasp the hand is able to perform intrinsic movement without having to move the arm; in the third category the intermediate grasp elements of power and precision grasp are present in the same proportion; with respect to the opposition types, three basic direction in which the hand can apply forces on the object to hold it securely, were mentioned: pad, palm and side opposition, in pad opposition category the object is parallel to the palm and hold between the fingers and the thumb (example holding a needle); palm position occurs when the object is hold using the finger, thumb and palm (example holding a hammer or screwdriver) and side position occurs when the object is transverse to the palm and is hold between the fingers and thumb (example holding a cigarette or a key). Concluding identifying 33 grasp types commonly used in the labor sector, arranged by the number of fingers in contact with the object and the position of the thumb.

1.2. Background and related work

The grasp type recognition has been implemented mainly in the area of robotics and automatic object manipulation. Zou et al. (2018) proposed a method for real-time grasp recognition using the leap motion controller (LMC) sensor, from which relative thumb positions, finger joint angles, and finger orientation were obtained and the used a Support Vector Machine (SVM) classifier to identify six types of grasps: large diameter, small diameter, index finger diameter, palmar pinch, prismatic and sphere. The accuracy of the classification model with the best performance is 75.65%. The authors concluded that the greatest confusion in prediction occurred in the small and large diameter grasp type, which they attribute to occlusion issues during motion capture.

Baldominos et al. (2019) compared the performance of machine learning techniques for human activity recognition using inertial sensors placed on the wrist and in the pocket of the analyzed subjects. Activities analyzed include walking, cycling, going up and down stairs, common postures (such as standing and sitting), work activities (such as writing and typing), and leisure activities (such as talking, eating, drinking coffee, and smoking). The results show that the higher accuracy and F-score (94.87% and 94.68%, respectively), were attained from the dataset obtained from the sensor placed on the wrist using the models trained through decision tree ensembles.

Moschetti et al. (2016) found that, for the prediction of activities of daily living that involve movements of the upper extremity, the accuracy and F-score obtained by support vector machine (SVM) classification models are higher than those obtained by models trained by decision trees, using one, two or three inertial sensors placed on the hand.

1.3. Activity Recognition Chain

Bulling et al. (2014) proposed a tutorial on human activity recognition using on-body inertial sensors, it provides an Activity Recognition Chain (ARC) as a general-purpose framework for designing, implementing and evaluating Human Activity Recognition (HAR) systems; the framework comprises components for data acquisition and pre-processing, data segmentation, features extraction and selection, training and classification, decision fusion and performance evaluation.

1.3.1 Sensor Data Acquisition and Preprocessing

In the first stage of a typical ARC, raw data is acquired using several sensors attached to different locations on the body (Bulling et al., 2014); preprocessing of acceleration and gyroscope signals may involve calibration, unit conversion, normalization, resampling, synchronization or signal level fusion (Figo et al., 2010).

1.3.2 Data Segmentation

Segmenting a continuous sensor stream is a challenging task since the exact boundaries of an activity are difficult to define. In the literature, various methods exist to approach the problem of segmentation the simplest is sliding window, where the sequence of n input samples is split into windows of w consecutive samples. However, the data associated with a particular event could be split in different windows, resulting in important loss information. An alternative segmentation method, frequently used for dividing the accelerometer data flow, is overlapping sliding window (Bulling et al., 2014; Figo et al., 2010; Lara & Labrador, 2013; Lima et al., 2019; Ni et al., 2016) , which includes overlap between adjacent windows, and the different overlapping percentages can be referred as step. The windows size, w, directly influences the delay of the system; the bigger window size, the longer the ARC has to wait for a new segment to be available for the processing, in the other hand, the larger the step size, the less frequently all subsequent stages of the ARC are executed which reduces the computational load.

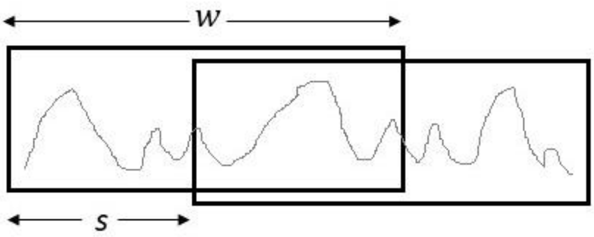
The Figure 1 shows the sliding window whit the fixed size of w samples and a step of s samples. Every time the windows slides over the data a new segment is created with the samples contained in the window. Each window i, covers an input vector x_i from the sequence of data T using Equation 1 where x_{start} is defined by Equation 2 (Baldominos et al., 2019).

$$x_i = T[x_{start}, x_{(start+w+1)}] \quad (1)$$

$$x_{start} = s * i \quad (2)$$

Figure 1.

Sliding window of size w and step s



Source: Adapted from Baldominos et. al. (2019).

Banos et al. (2014) presented an extensive study to characterize the window procedure to determine its impact within the activity recognition process, demonstrating that the most precise recognizer is obtained for very short windows (0.5 to 1 second for arms activities) contrary to what its often thought, large windows sizes do not necessarily translate into a better recognition performance. Achumba et al. (2012) determined the optimal segmentation approach using 12 second windows size and 90 % windows overlap to partition measured human ambulation and activity sensor data.

Is clear that decreasing the windows size allows for a faster activity detection, but large data windows are preferred for complexity activities; in the other hand very, small step can lead an unmanageable volume of data. The choice of segmentation techniques is in general very important because inappropriate segmentation will most likely result in features without discriminant power.

1.3.3 Feature Extraction and Selection

Feature engineering involves the selection and extraction of features. The first one consists of manually choosing the variables that will be considered in the HAR and is generally done based on experience and prior knowledge about the HAR. On the other hand, the extraction of characteristics consists of the transformation process of the data contained in each one of the windows defined in the segmentation stage; this process seeks to determine a set of features that minimizes classification errors and reduces computation time (Nweke et al., 2018). Feature extraction is a necessary step in the analysis of signals from inertial sensors, since the raw data obtained by such sensors are not suitable for use in machine learning algorithms.

Various studies have analyzed feature selection and its impact on the performance of machine learning classification models trained for RAH, Shoaib et al. (2014) indicate that the accuracy of the prediction can be increased if the information coming from two sensors (usually accelerometer and gyroscope) are combined, at the same time that the performance for different subjects is stabilized. In Lima et al. (2019) and Chen & Shen (2017) it's also suggested that each sensor has different abilities to discriminate activities, so the selection of sensors depends on the specific task under analysis.

Regarding the feature extraction process, the literature classifies them into different representation domains: frequency domain, time domain, and discrete domain. In each domain there are different mathematical models that allow extracting specific characteristics of the signals. In the time domain, data can be represented by statistical functions: minimum, maximum, mean, standard deviation, and non-statistical functions: area and frequency histograms. The most commonly used characteristics in the time domain for the HAR are the mean and the standard deviation (Bulling et al., 2014; Figo et al., 2010; Moschetti et al., 2016; Shoaib et al., 2014). On the other hand, in the frequency domain it is possible to carry out an analysis of the signal based on the frequency spectrum of the values of a certain data window, which allows the identification of repetitive patterns of the signals related to the natural periodicity of the signals activities, Fast Fourier Transform (FFT) and Wavelet characteristics are commonly used in this domain. In Baldominos et al. (2019), it is recommended to obtain the median, lower and upper quartile, the coefficient of kurtosis and skewness in the HAR. The extraction of features in the time domain is more widely used than the extraction through the frequency domain, due to the lower computational cost that it represents in the analysis of signals from inertial sensors.

1.3.4 Training and Classification

Machine learning is a branch of artificial intelligence that seeks to build mathematical models in order to make predictions, in this sense, learning is defined as the ability to learn according to an external stimulus and remember most of the previous experiences (Chandra Sen et al., 2020; Jianwei Niu et al., 2010).

In the literature there are several traditional strategies that have been used in HAR, requiring feature engineering, which consists of selecting the most relevant dimensions or features to be used in training classification models, such as k-nearest neighbors (kNN), support vector machines (SVM), decision trees (DT) and ensembles (ENS) (LeCun et al., 2015). Traditional classification methods include distance, statistical, kernel and decision tree based methods, also ensemble learners (Saez et al., 2016). Traditional machine learning methods can be classified based on the characteristics of the data set used: supervised learning and unsupervised learning. In supervised learning, models are trained that allow members of a data set to be classified based on a previous label or classification, while new members obtain a prediction based on the learning of the model. In the case of unsupervised learning, the models are trained without knowing the label or the class to which they belong, allowing to identify the natural patterns (groups) of the data set (LeCun et al., 2015).

Statistical methods:
These techniques assume that the data follow a probability function. Probabilistic models perform learning by inferring the probability distribution of data. The most commonly used method is the Naive-Bayes (NB) algorithm, which assumes strong independence between the predictors. The algorithm makes the prediction according to the class that provides a higher conditional probability (Saez et al., 2016).

Kernel-based methods:
These methods perform pattern analysis based on a kernel function, which is the similarity function between data pairs. The Support Vector Machine (SVM) algorithm is commonly used in classification and regression problems. Some of the advantages of this method is that it has a high performance in high-dimensionality data sets, in addition to providing a solution with global optimum characteristics, allowing the generalization of the results (Xue et al., 2019). The SVM algorithm was designed for binary classification, however, there exist strategies by which they can be adopted to multiclass tasks associated with remote sensing studies through decomposition into binary analysis series, which are treated as binary SVM problems, combining one-against-one and one-against-all strategies (Anthony et al., 2007). The one-against-all strategy divides the data set of N classes, into N cases of two classes. In contrast, the one-against-one strategy builds a machine for each pair of classes resulting in

$(N(N\text{-}1))\,/\,2$ machines. In the test stage with the test set, each classification gives one vote to the winning class and the observation is tagged with the class containing the most votes (Xue et al., 2019). One of the main disadvantages of the one-against-one strategy is that it requires more computational intensity, whereas the performance of the one-against-all strategy may be lower on an unbalanced data set. If the classes representing the hand movements are not linearly separable, kernel selection is crucial for good classifier performance.

Ensemble Learners:
These models combine the predictions of several base estimators, built using a given learning algorithm, in order to improve the results and the robustness of a simple estimator *(Saez* et al., 2016). The idea behind of ensembles is that each classification model is specialized in a set of instances, and when working together they perform majority voting in order to decide which class should be assigned to a certain instance.

1.3.5 Decision Fusion

The decision fusion stage combines several intermediate classification results into a single decision. Fusion can happen at features or classifiers stages, some rules used in activity recognition are summation, majority voting, and Bayesian fusion, some sensor fusion benefits for an HAR system are: increased robustness, reduced classification problem complexity through use of classifiers dedicated to a specific subsets, classification with missing features and discriminative training (Bulling et al., 2014).

1.3.6 Performance Evaluation

The performance of the classification models is evaluated through metrics that objectively indicate the reliability of the model in the HAR process (Lima et al., 2019), such as: accuracy, sensitivity, specificity, precision, recall and F-score (Bulling et al., 2014; Lara & Labrador, 2013). The mentioned metrics can be obtained through the information provided by the confusion matrix M_{nxn} for a classification problem with n classes, confusion matrix associates the predictions of the class to which a member of the data set belongs, and the current classification (label or class that was used in the training phase of the model). Table 1 shows the confusion matrix in a binary classification problem (n=2).

Table 1.

Confusion matrix for n=2**.**

	Negative Prediction	**Positive Prediction**
True Negatives	Number of negatives instances that were classified as negative, True Negative (TN)	Number of negatives instances that were classified as negative, False Negative (FN)
True Positives	Number of positive instances that were classified as negative, false negative (FN)	Number of positive instances that were classified as positive, True Positive (TP)

Source: Lara & Labrador (2013).

The precision indicates, often referred as a positive predictive value, is the proportion of correctly classified positive instances, with respect to the total number of real instances of the class (equation 3). Recall measures, also called true positive rate is the proportion of positive instances correctly classified, with respect to the total number of instances predicted for the class (equation 4). The accuracy metrics (equation 5) and the F-score (equation 6) are global measures, the first one denotes the global performance of the classifier model for all classes, while the F-score relates precision and recall (Lara & Labrador, 2013).

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

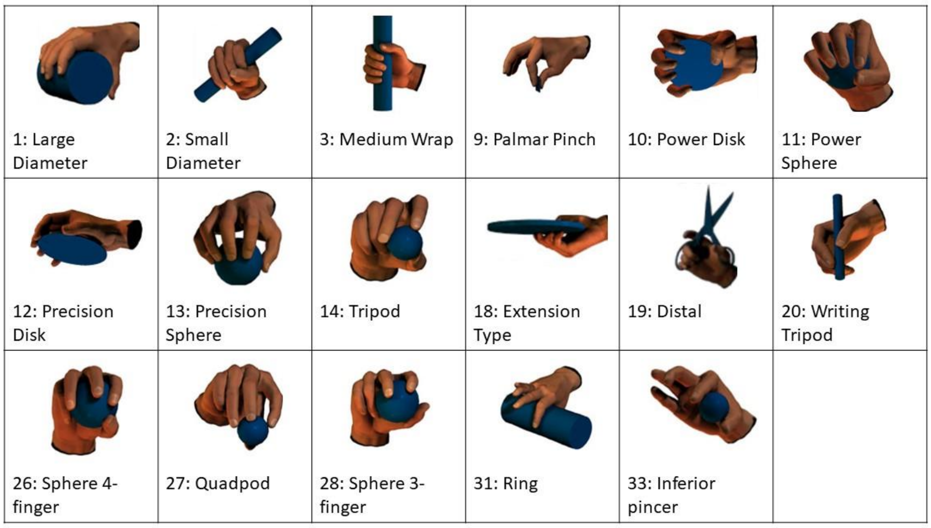
$$F - Score = 2 * \frac{Precision * Recall}{Precision+Recall} \quad (6)$$

2. Methodology

The GRASP taxonomy, which allows classifying the grasp commonly used in the labor sector into 33 types (Feix, et al., 2016) is the base of this study, seventeen of the 33 types of grasps, including power and precision grasp, were selected with the thumb in the two positions considered in the taxonomy: abduction and adduction (Figure 2). In order to identify different movement patterns, objects with different characteristics were used for the types of grasps studied.

Figure 2.

Grasp types selected for identification

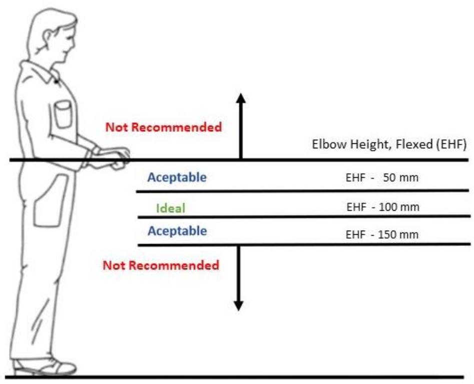


Source: Adapted from Feix et al. (2016).

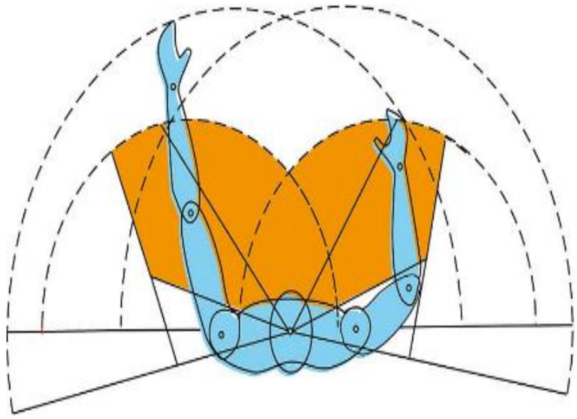
The experiment was carried out in laboratory conditions, in which a work station was simulated, the objects were positioned on a workbench placed in front of the individual with working heights recommended for standing work, presented in Figure 3(a),within the optimal working area for industrial assembly activities, orange zone, showed in Figure 3(b) (Anthony et al., 2007).

Figure 3.

Experiment workstation description (a) Working heights recommended for standing work, (b) Optimal working area for industrial assembly activities



(a)



(b)

Source: Adapted from Anthony et al. (2007)

Participants were instructed on the correct way to perform the different types of grasps considered in the study. At the time of recording the data, they were labeled in order to identify the start and end of each type of movement. At the beginning of the recording of measurements of each movement, each participant was asked to position the elbow flexed at 90 degrees, to start the movement in the neutral posture in the upper extremity.

The analysis of the data obtained was carried out according to the HAR methodology indicated in (Bulling et al., 2014), which is described in the following sections.

2.1. Sensor Data Acquisition and Preprocessing

To carry out the registration of information corresponding to the manual manipulation of objects, a data glove made up of inertial sensors and resistive force sensors was used; five inertial sensors with nine degrees of freedom were placed on the proximal phalanges of each finger and one more on the back and in addition, five resistive force sensors were placed on each fingertip and one more on the palm hand. Acceleration, angular velocity and magnetic components are obtained by inertial sensors, showing information about the hand and fingers movement while contact forces are obtained by the forces resistive sensors showing information about the contact of the hand with the manipulated object. The Figure 4 shows the sensors position. The sampling frequency is 25 Hz.

The data obtained through the data glove were acquired and processed using MATLAB® version 2019[a] installed on a laptop computer with a 64-bit Windows 10 operating system with Intel Core i7-8550U CPU and 16 GB of RAM. The motion capture system was connected to the laptop computer through a USB port.

Figure 4.

Designed data glove for the experiment (a) Force Resistive Sensors in the palm hand (b) Inertial Sensors in the back hand.

(a) (b)

Source: Developed by authors.

The raw data were sorted chronologically in a tridimensional array for the acceleration, angular velocity and magnetic components. The zero motion and zero rate methods for accelerometer and gyroscope calibration were used; zero motion method consists of obtaining the average of a thousand accelerometer readings, such readings are taken when the sensor is at rest, zero motion, on a flat surface and with the z axis oriented in the opposite direction to the flat surface, while the zero rate method consists of calculating the average of 1000 gyroscope readings while the sensor is at rest, zero rate, regardless of its orientation. On the other hand, the calibration of the magnetometer includes the calculation of the hard (hard iron) and soft (soft iron) deviations of the magnetic fields according to Lima et al. (2019), implemented using the *magcal function* in MatLab 2019[a]. Subsequently, a fifth-order median filter was used to remove abnormalities and peaks in the acquired signals that can be caused by environmental conditions, movements and behavior changes of the user (Lima et al., 2019).

2.2. Data Segmentation

There is evidence in the literature that segmentation with overlapping fixed-size windows handles transitions more precisely than those that do not share information (Lapucci et al., 2021) . Some HAR studies have been performed with 50% overlap, while others recommend a fixed window size of 2 to 5 seconds, considering a sampling frequency of 20Hz to 50Hz (Banos et al., 2014; Feix et al., 2014; Kubota et al., 2019; Shoaib et al., 2014). In the present study, different window size combinations were selected with two overlap options: 50% and greater than 50% as shown in Table 2.

Table 2.

Description of sliding windows and step size combination for data segmentation.

Code	Windows size		Step Size		Overlap
	Samples	Seconds	Samples	Seconds	
W50 S25	50	2	25	1	50 %
W75 S35	75	3	35	1.4	50 %
W100 S50	100	4	50	2	50 %
W25S10	25	1	10	0.4	>50 %
W50 S10	50	2	10	0.4	>50%
W75 S10	75	3	10	0.4	>50%
W100 S10	100	4	10	0.4	>50%

Source: Developed by authors

The time series data were transformed to the frequency domain using the fast Fourier transform available in MatLab 2019[a] and only the real coefficients were considered, ignoring the imaginary part of the transformed values as in Baldominos et al. (2019). The following characteristics were extracted: mean and standard deviation of the raw data vector, as well as median, low and upper quartile, and coefficients of skewness and kurtosis of the transformed vector.

2.3. *Trainning and Classification*

Traditional classification models were trained: support vector machines (SVM), ensembles (ENS), Naive-Bayes (NB) algorithm, cross-validation with k=10 was used to avoid overtraining.

3. Results

Tables 3, 4 and 5 show the accuracy values and F-score corresponding to the models trained for the different window-step combinations. Both metrics indicate that the models trained through support vector machine, (SVM) have a better performance, followed by the models corresponding to ensembles (ENS), while the models trained through Naive-Bayes (NB) presented greater confusion in the prediction of the type of grip.

Regarding the features selection, considering only the components of the acceleration and force sensors (table 3), or the acceleration, angular velocity and magnetics components of the readings coming from the inertial sensors in addition to the force sensors (table 5), presented similar results, but of lower magnitude than the predictions obtained by considering the acceleration and angular velocity from the inertial sensors plus the force resistive sensors (table 4).

The W75S10 segmentation (window size=75 samples, step=10 samples) presented the highest value F-Score, for the different feature selection options.

Table 3.

Trained models accuracy and F-Score using acceleration components and force sensors

Window-step	Accuracy			F-score		
	ENS	SVM	NB	ENS	SVM	NB
W25 S10	0.7984	**0.8564**	0.3748	0.7967	**0.8483**	0.3990
W50 S10	0.8503	**0.8718**	0.4105	0.8481	**0.8649**	0.4369
W75 S10	0.7821	**0.8748**	0.4126	0.7805	**0.8711**	0.4504
W100 S10	0.7607	**0.8644**	0.4763	0.7472	**0.8623**	0.4196
W50 S25	0.7109	**0.8355**	0.3010	0.7024	**0.8299**	0.3164
W70 S35	0.6385	**0.8215**	0.2846	0.6041	**0.8187**	0.3161
W100 S50	0.5399	**0.7073**	0.2175	0.4929	**0.7226**	0.2338

Source: Developed by authors based on data.

Table 4.

Trained models accuracy and F-Score using acceleration and angular velocity components plus force resistive sensors

Window-step	Accuracy			F-score		
	ENS	SVM	NB	ENS	SVM	NB
W25 S10	0.8405	**0.8687**	0.3911	0.8344	**0.8620**	0.4186
W50 S10	0.8090	**0.8964**	0.4258	0.7998	**0.8921**	0.4558
W75 S10	0.8570	**0.9066**	0.4246	0.8537	**0.9047**	0.4626
W100 S10	0.8491	**0.9007**	0.4039	0.8387	**0.8940**	0.4383
W50 S25	0.8198	**0.8700**	0.3118	0.8044	**0.8671**	0.3291
W70 S35	0.8469	**0.8669**	0.2931	0.8357	**0.8618**	0.3244
W100 S50	0.7836	**0.7973**	0.2232	0.7686	**0.7985**	0.2423

Source: Developed by authors based on data.

Table 5.

Trained models accuracy and F-Score using acceleration, angular velocity and magnetic components plus force resistive sensors

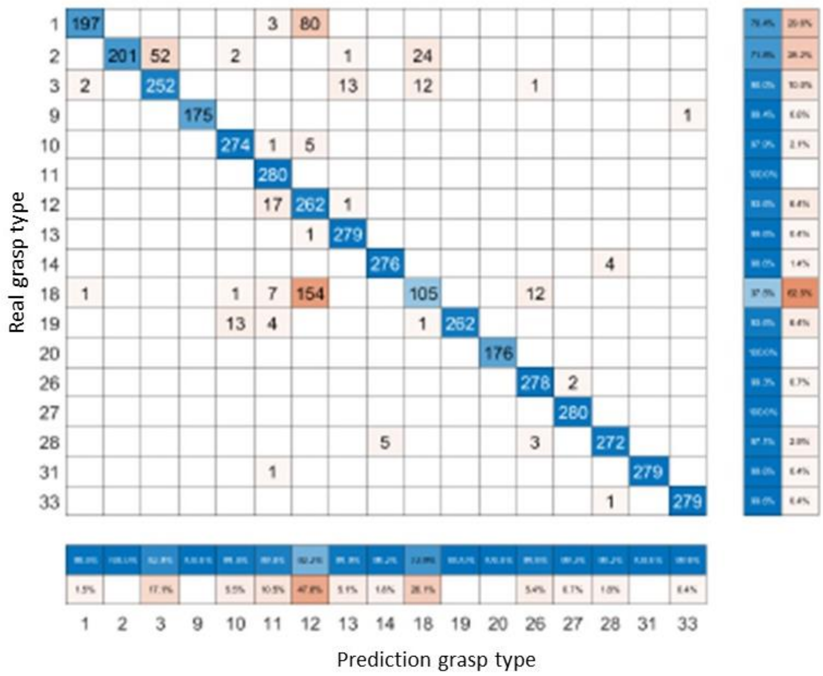
Window-step	Accuracy			F-score		
	ENS	SVM	NB	ENS	SVM	NB
W25 S10	0.8463	**0.8530**	0.3676	0.8354	**0.8484**	0.3974
W50 S10	0.8609	**0.8754**	0.4273	0.8557	**0.8719**	0.4576
W75 S10	0.8366	**0.8801**	0.4200	0.8291	**0.8769**	0.4568
W100 S10	0.8406	**0.8792**	0.4091	0.8385	**0.8767**	0.4432
W50 S25	0.7983	**0.8371**	0.3080	0.7897	**0.8331**	0.3276
W70 S35	0.8131	**0.8362**	0.3008	0.8081	**0.8319**	0.3318
W100 S50	0.5900	**0.7984**	0.2323	0.5787	**0.7937**	0.2508

Source: Developed by authors based on data.

Figure 5 presents the confusion matrix for the best results combination: SVM classification model, acceleration and angular velocity components plus force resistive sensors, W75S10 combination (window size 75 samples and step size 10 samples). This matrix shows the frequencies of the correct predictions (values on the diagonal of the matrix, in blue) as well as the predictions that presented confusion considered incorrect (outside of the diagonal values, showed in brown). From the aforementioned confusion matrices, it can be concluded that the most confused grips correspond to type 1, 2 and 18 grips, which are predicted as type 12, 3 and 12 grips, respectively.

Figure 5.

Confusion Matrix for the SMV model using the acceleration and angular velocity components plus force resistive sensors with the combination of window and step size W75S10.



Source: Developed by authors based on data.

4. Discussion

Figure 6 shows the comparation between the real grasp type and the predicted. The confusion in the prediction of grasp type 1, being classified as grip type 12 may be due to the nature of the movement, this is because the manipulation movements of the objects were carried out in different positions of the hand-wrist: pronation, supination and neutral posture. In the case of the prediction of grasp type 2 as grasp type 3, it could be caused by the similarity of the movement, when using objects of similar diameter. The confusion in the prediction of grasp type 18, being classified as grasp type 12, can be attributed to the physical characteristics of the manipulated object (extreme diameter and minimum weight) and the size of the hands of the subjects who carried out the experimentation, forcing a power grip of such width that the fingers remain with minimal flexion.

In relation to similar studies, it is observed that in Baldominos et al. (2019) accuracy and F-score of 94.87% and 94.68% respectively are obtained using models trained by ensembles, (ENS) using features obtained by inertial sensors placed at two mobile devices, however, the activities identified in this study are full-body, unlike those proposed in the present study. On the other hand, the results obtained corroborate the results presented in Moschetti et al. (2016), in which it is concluded that the models trained using SVM have a higher accuracy and F-score than those obtained using decision trees. A better performance is observed in this study, than the one presented in Zou et al. (2019) through the combination of inertial sensors and resistive force sensors, (75.65%), who used the Leap Motion Controller sensor and SVM. The window size that presented optimal performance in the present study (W75S10, which represents a window size = 3 seconds with 75 samples), is higher than the one suggested in Banos et al. (2014) whose range goes from 0.75 to 2.25 seconds.

The present study has two main limitations. The first one refers to the controlled laboratory conditions under which the experimentation was carried out. The second is related to the discriminatory power of the classification models for the prediction of the grasp type, which is influenced by the configuration of the sensors used in the data glove used, in addition to the objects handled. It is suggested as future work to carry out the analysis with variations in the number of inertial and resistive sensors included in the data glove, as well as the placement of these in different points of the fingers and hands.

Figure 6.

Grasps type real and predicted comparation.

Real Grasp Type	Predicted Grasp Type
 1. Large Diameter	 12. Precision Disk
 2. Small diameter	 3. Medium wrap
 18. Extension type	 12. Precision Disk

Source: Developed by authors based on data.

5. Conclusions

The results indicate that data segmentation in sliding windows with overlap greater than 50%, are more recommendable than the segmentation in sliding windows with overlap less than or equal to 50%. On the other hand, the grasp types that presented the greatest confusion in the prediction were grips 18, 2, and 1, being classified as grips 12, 3, and 1, respectively.

In the case of the features selection, it is possible to infer that under the conditions in which the sampling of activities was carried out, the grasp type recognition by traditional classification models presents a smaller error in the prediction when considering the features combination of acceleration and angular velocity components from inertial sensors, plus readings from force resistive sensors; this can be generalized for conditions similar to those considered in the present study.

The limitations of this study make it possible to identify opportunities for future work, in which more inertial and force resistive sensors are included that allow a more precise discrimination between the grasps type, in addition to the analysis of the manipulation of objects with features not considered in this study plus to considering experimental environments that are more representative of real working conditions.

6. References

Achumba, I. E., Bersch, S., Khusainov, R., Azzi, D., & Kamalu, U. (2012). *Activity Classification*, 427-430.

Anthony, G., Gregg, H., & Tshilidzi, M. (2007). Image classification using SVMs: One-Against-One Vs One-against-All. *28th Asian Conference on Remote Sensing 2007, ACRS 2007*, 2, 801-806.

Armstrong, T. J., Foulke, J. A., Joseph, B. S., & Goldstein, S. A. (1982). Investigation of cumulative trauma disorders in a poultry processing plant. *American Industrial Hygiene Association Journal*, *43*(2), 103-116. https://doi.org/10.1080/15298668291409433

Baldominos, A., Cervantes, A., Saez, Y., & Isasi, P. (2019). A comparison of machine learning and deep learning techniques for activity recognition using mobile devices. *Sensors (Switzerland)*, *19*(3). https://doi.org/10.3390/s19030521

Banos, O., Galvez, J. M., Damas, M., Pomares, H., & Rojas, I. (2014). Window size impact in human activity recognition. *Sensors (Switzerland)*, *14*(4), 6474-6499. https://doi.org/10.3390/s140406474

Barr, A. E., Barbe, M. F., & Clark, B. D. (2004). Work-Related Musculoskeletal Disorders of the Hand and Wrist: Epidemiology, Pathophysiology, and Sensorimotor Changes. *Journal of Orthopaedie & Sports Physical Therapy*, *34*(10), 610-627.

Bevan, S. (2015). Economic impact of musculoskeletal disorders (MSDs) on work in Europe. *Best Practice and Research: Clinical Rheumatology*, *29*(3), 356-373. https://doi.org/10.1016/j.berh.2015.08.002

Bulling, A., Blanke, U. L. F., & Schiele, B. (2014). *A Tutorial on Human Activity Recognition Using Body-Worn*, *46*(3), 1-33.

Chandra Sen, P., Hajra, M., & Ghosh, M. (2020). Supervised Classification Algorithms in Machine Learning: A Survey and Review. In J. Mandal, and D. Bhattacharya (*Eds*.). *Emerging Technology in Modelling and Graphics. Advances in Intelligent Systems and Computing* (pp. 99-111). Springer Singapore. https://doi.org/10.1007/978-981-13-7403-6

Chen, Y., & Shen, C. (2017). Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition. *IEEE Access*, *5*(c), 3095-3110. https://doi.org/10.1109/ACCESS.2017.2676168

Feix, T., Bullock, I. M., & Dollar, A. M. (2014). Analysis of human grasping behavior: Object characteristics and grasp type. *IEEE Transactions on Haptics*, *7*(3), 311-323. https://doi.org/10.1109/TOH.2014.2326871

Feix, T., Romero, J., Schmiedmayer, H. B., Dollar, A. M., & Kragic, D. (2016). The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions on Human-Machine Systems*, *46*(1), 66-77. https://doi.org/10.1109/THMS.2015.2470657

Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, M. P. (2010). Preprocessing techniques for context recognition from accelerometer data. *Pers Ubiquit Comput*, 14, 645-662. https://doi.org/10.1007/s00779-010-0293-9

Govaerts, R., Tassignon, B., Ghillebert, J., Serrien, B., De Bock, S., Ampe, T., El Makrini, I., Vanderborght, B., Meeusen, R., & De Pauw, K. (2021). Prevalence and incidence of work-related musculoskeletal disorders in secondary industries of 21st century Europe: a systematic review and meta-analysis. *BMC Musculoskeletal Disorders*, *22*(1), 1-30. https://doi.org/10.1186/s12891-021-04615-9

Jianwei, N., Yiling, H., Muyuan, L., Xin, Z., Linghua, R., Chuzhi, Ch., & Baoqin, Z. (2010). A comparative study on application of data mining technique in human shape clustering: Principal component analysis vs. Factor analysis. *2010 5th IEEE Conference on Industrial Electronics and Applications*, 2014-2018. https://doi.org/10.1109/ICIEA.2010.5515577

Kubota, A., Iqbal, T., Shah, J. A., & Riek, L. D. (2019). Activity recognition in manufacturing: The roles of motion capture and sEMG+inertial wearables in detecting fine vs gross motion. *Proceedings - IEEE International Conference on Robotics and Automation*, 6533-6539. https://doi.org/10.1109/ICRA.2019.8793954

Lapucci, T., Troiano, L., Carobbi, C., & Capineri, L. (2021). Soft and hard iron compensation for the compasses of an operational towed hydrophone array without sensor motion by a helmholtz coil. *Sensors*, *21*(23), 8104. https://doi.org/10.3390/s21238104

Lara, D., & Labrador, M. A. (2013). A Survey on Human Activity Recognition using. I*EEE Communications Surveys & Tutorials*, *15*(3), 1192-1209. https://doi.org/10.1109/SURV.2012.110112.00192

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. https://doi.org/10.1038/nature14539

Lima, W. S., Souto, E., El-khatib, K., Jalali, R., & Gama, J. (2019). Human Activity Recognition Using Inertial Sensors in a Smartphone : An Overview. *Sensors, 19*(14), 3213; https://doi.org/10.3390/s19143213

Moschetti, A., Fiorini, L., Esposito, D., Dario, P., & Cavallo, F. (2016). Recognition of daily gestures with wearable inertial rings and bracelets. *Sensors, 16*(8), 1341. https://doi.org/10.3390/s16081341

Ni, Q., Patterson, T., Cleland, I., & Nugent, C. (2016). Dynamic detection of window starting positions and its implementation within an activity recognition framework. *Journal of Biomedical Informatics*, 62, 171-180. https://doi.org/10.1016/j.jbi.2016.07.005

Nur, N. M., Dawal, S. Z., & Dahari, M. (2014). *The Prevalence of Work Related Musculoskeletal* Disorders Among Workers Performing Industrial Repetitive Tasks in the Automotive Manufacturing Companies. In *Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management Bali*, 1-8. http://ieomsociety.org/ieom2014/pdfs/303.pdf

Nweke, H. F., Teh, Y. W., Al-garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, 105, 233-261. https://doi.org/10.1016/j.eswa.2018.03.056

Saez, Y., Baldominos, A., & Isasi, P. (2016). A Comparison Study of Classifier Algorithms for Cross-Person Physical Activity Recognition. *Sensors (Basel, Switzerland)*, *17*(1), 66. https://doi.org/10.3390/s17010066

Shoaib, M., Bosch, S., Durmaz Incel, O., Scholten, H., & Havinga, P. J. M. (2014). Fusion of smartphone motion sensors for physical activity recognition. *Sensors, 14*(6). https://doi.org/10.3390/s140610146

Xue, Y., Ju, Z., Xiang, K., Chen, J., & Liu, H. (2019). Multimodal Human Hand Motion Sensing and Analysis-A Review. *IEEE Transactions on Cognitive and Developmental Systems*, *11*(2), 162-175. https://doi.org/10.1109/TCDS.2018.2800167

Zou, Y, Liu, H. & Zhang, J. (2019). Real-Time Grasp Type Recognition Using Leap Motion Controller. *In: Yu, H., Liu, J., Liu, L., Ju, Z., Liu, Y., Zhou, D. (Eds.). Intelligent Robotics and Applications. ICIRA 2019. Lecture Notes in Computer Science* (pp.472-480). https://doi.org/10.1007/978-3-030-27535-8_42

AUTHORS' CONTRIBUTIONS, FINANCING AND ACKNOWLEDGMENTS

Contributions of the Authors:

Conceptualization: Rodríguez-Vega, G.; **Software:** Rodríguez-Vega, G.; **Validation:** Rodríguez-Vega, D.A. **AFormal Analysis:** Rodríguez-Vega, G.; **Data curation:** Rodríguez-Vega, D.A.**;** **Writing-Preparation of the original draft:** Rodríguez-Vega, D.A. **Drafting-Revision and Editing:** Rodríguez-Vega, D.A. **Visualization:** Rodríguez-Vega, D.A. **Supervision:** Rodríguez-Vega, G. **Project Administration:** Rodríguez-Vega, D.A. **All Authors have read and accepted the published version of the manuscript:** Rodríguez-Vega, G., Rodríguez-Vega, D.A.

AUTHOR:

Graciela Rodríguez Vega
Universidad de Sonora, México.

Industrial Engineer by the TecNM/Instituto Tecnológico de Los Mochis. Master of Science in Industrial Engineering from the TecNM/Instituto Tecnológico de Hermosillo. PhD in Information Sciences from the Universidad Autónoma de Sinaloa. She is a Full-Time Research Professor at the University of Sonora and a member of the National System of Researchers in the candidate category. She has worked in the area of Occupational Health, mainly in Ergonomics and Anthropometry, as well as in Machine Learning and Ergonomic Risk Level Prediction.
graciela.rodriguez@unison.mx

Orcid ID: https://orcid.org/0000-0001-7366-1107

Dora Aydee Rodríguez Vega
Universidad Politécnica de Sinaloa, México.

Dora Aydee Rodríguez Vega has a degree in Engineering, a Master of Science in Electronic Engineering with a specialisation in Image Processing and a PhD in Information Sciences. She has worked as a Full-Time Professor in the Mechatronics Engineering programme at the Polytechnic University of Sinaloa since 2005. She has a desirable PRODEP profile since 2007, is a member of the Sinaloan System of Researchers and Technologists in the Researcher category and a member of the National System of Researchers in the candidate category. Her line of research is the imitation of movements in humanoid robots, which includes the capture and interpretation of movements and their reproduction in humanoid robots.
drodriguez@upsin.edu.mx

Índice H: 3
Orcid ID: https://orcid.org/0000-0002-6521-7978
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57222389070
Google Scholar: https://scholar.google.es/citations?hl=es&pli=1&user=tXjq6XAAAAAJ