Multimedia Tools and Applications (2023) 82:23833–23871
https://doi.org/10.1007/s11042-022-14117-y

TRACK 3: BIOMETRICS AND HCI

A reliable and efficient machine learning pipeline for american sign language gesture recognition using EMG sensors

Shashank Kumar Singh[1] · **Amrita Chaturvedi**[1]

Received: 25 January 2022 / Revised: 14 June 2022 / Accepted: 25 October 2022 /
Published online: 17 November 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Sign languages has extensive applications among differently-abled to communicate with their surroundings. With the development of different sensing technologies, several new human-computer interaction techniques (HCI) have been established to recognize hand gestures. Computer vision-based methods have shown significant utility for such applications. However, these methods are strongly dependent on the lighting conditions. The surface electromyography (sEMG) technique is invariant to lighting conditions and can easily reflect human motion intention. In this work, sEMG based sign language recognition model was developed using an efficient machine learning pipeline.

Two sEMG datasets were recorded for predefined hand gestures using wireless sensors. These signals were mainly acquired against 24 manual alphabets (ASL-24) and ten digits(ASL-10) of American Sign Language (ASL). The collected data sets were preprocessed, and around 450 well-established feature was extracted from each sEMG channel. We applied an ensemble feature selection approach combining four diverse filter-based feature selection methods (ANOVA, Chi-square, Mutual Info, ReliefF). A newly proposed feature combiner that exploits feature–feature and feature–class correlation thresholds is used to combine feature subsets formed across the ensemble. The resulting features comprise reduced & most representative feature subsets and are further used in the pipeline for classifying ASL gestures.

Using the CatBoost algorithm, the pipeline presented excellent average classification accuracy(99.91% on ASL-24) and other performance parameters for recognizing ASL gestures. The pipeline was also applied and validated on a benchmark dataset (Ninapro database 5, exercise A) and achieved similar outcomes. The result highlights the feasibility of using sEMG based approach as better options to computer-vision-based techniques to build an accurate and robust Sign Language Recognition system (SLRS). Moreover, efforts were made to find the optimal number of sensors and features for recognition task on (ASL-10 dataset) without impacting the overall reliability and accuracy of the system. The experiments results can be used to enhance the performance of various wearable sEMG sensor based HCI applications.

✉ Shashank Kumar Singh
shashankkrs.rs.cse17@itbhu.ac.in

Extended author information available on the last page of the article.

Keywords Human–computer interaction · Surface electromyography · American sign language · Feature selection

1 Introduction

Sign languages are considered as visual and non-verbal form of communication used by differently-abled people to express themselves or interact with their surroundings. These languages are expressed using manual and non-manual features. These features mainly include different hand movements, use of different number of fingers, palm orientation, facial expression, head orientation, hand shape, eye gaze, etc. [19, 57]. American Sign Language (ASL) is one of the most widely used sign languages that consists of various static and dynamic gestures as a means of human expression [89]. Most of these gestures are generated by a single-hand [109]. ASL is similar to any other natural language, having syntax and grammatical rules of its own [9]. However, there are a few physical disabilities in some people due to which it becomes difficult for them to convey and/or understand the meaning of sign language completely [74]. Also, in the absence of global standardization, there are many different regional sign languages that widely differ in their nature, lexicon grammar, and style [65]. So, it becomes a complicated task for individuals from diverse regions with different sign languages to communicate with each other. A computer-assisted Sign Language Recognition system (SLRS) can be quite helpful in solving such problems. The computer-based recognition system can assist, automate, act as a translator, and facilitate communication for differently-abled persons [90].

Advancements in modern wearable sensing technologies and novel human-computer interaction techniques (HCI) assisted with machine learning have enhanced the reliability and efficiency of these SLRS systems. In most of the research works ASL recognition is performed using different sensing devices such as Kinetic sensors [76, 78], Hand Gloves [104], Leap motion controller (LED Motion sensors and cameras) [18], Surface Electromyography(sEMG) sensors [85, 86, 92], Inertial Sensors, Pressure sensors [56] and use of 2D/3D cameras (image and video) [7, 32, 43, 79]. Based on data input mechanism, these approaches are classified as Visual, Sensor-based, and Hybrid methods [95]. For Visual SLRS, the input data consists of images or video streams. However, Sensor-based SLRS utilizes input data from various sensors such as datagloves, sEMG sensors, accelerometers, etc. The hybrid SLRS approach benefits from the combination of elements of both approaches.

Compared to other approaches, Visual sign language recognition systems have been extensively used to build SLRS [95]. However, the efficiency of these methods depends on environmental lighting conditions and requires specific infrastructure. These requirements are hard to maintain in standard conditions in a real-world scenario. On the other hand, Sensor-based SLRS are independent or less affected by lighting conditions and are cheaper than vision-based SLRS. Of the entire research conducted for ASL recognition system, only 21% work has been done using wearable sensors or gloves [95]. This fact motivated us to analyze, explore, and use the sEMG signals for sensor-based ASL recognition model. EMG signals are bio-signals that represent the electrical activity of the muscles while a person is active or is having rest [17]. Being non invasive in nature, sEMG signals have wide applications in various fields including prosthetic control, gesture recognition, HCI, and Rehabilitation [21, 83]. The availability of sEMG sensors as wearable wireless modules and being invariant to lighting conditions, makes these sensors preferable candidates for building efficient SLRS. However, to achieve higher recognition accuracy, the researchers

applied multiple sEMG sensors and even fused them with additional inertial sensors. Savur and Sahin [53, 85, 86].

Despite the developments, a cost-effective SLRS is not available commercially [5]. The cost and efficiency of these SLRS depend on the number of sensors used. To build a cost-effective SLRS, a trade-off between these factors must be maintained. We intend to fill this gap by proposing a methodology to build a low cost SLRS that uses an optimal number of sEMG sensors and minimum number of channels yet provides a better performance as compared to existing works.This paper explores the feasibility of building a reliable and accurate sign language recognition model using wearable sEMG sensors. For the same we proposed a machine learning pipeline that composed of a new ensemble features selection method, exhaustive set of extracted sEMG features, and an a efficient boosting algorithm for classifying the ASL gestures. Other objective of this work is to provide an interpretable Machine learning pipeline in terms of features used and with improved or comparable classification accuracy with the state-of-art methods. Also, an effort is made to use optimal number of sEMG sensors by reducing the number of sEMG channels.

We created two ASL gestures dataset(ASL-10 & ASL-24) by collecting raw sEMG signals from various people(up-to 20 subjects).ASL-10 dataset consist of sEMG signals collected for 10 ASL digit gestures. While the ASL-24 consists of sEMG signals collected for 24 manual alphabet of the ASL dictionary. Around 450 time and frequency domain features were extracted from each sEMG channel. We applied an ensemble feature selection approach combining four diverse filter-based feature selection methods (ANOVA, Chi-square, Mutual Info, ReliefF). Initially, k-best features are extracted from filter-based methods using the feature ranking method. Then, a greedy search approach is used to select subsets from these various filter-based approaches based on the ranking of features and feature importance heuristic. Finally, a newly proposed feature combiner that exploits feature–feature and feature–class correlation thresholds is used to combine feature subsets formed across the ensemble. The resulting features comprised reduced and most representative feature subsets for classifying ASL gestures.

The pipeline achieved excellent average classification accuracy of 99.99% (ASL-10) and 99.91% (ASL-24) on the two aggregated datasets, using two and eight SEMG channels, respectively. Moreover, the machine learning pipeline and the feature set obtained from ensemble feature selection from (ASL-10 datset) were validated on a subset of the benchmark dataset (Ninapro database 5), displaying similar results. The Fast Fourier Transform coefficients dominated the selected features set as compared to other extracted features for the classification task. The outline of our proposed pipeline is illustrated in Fig. 1. The major contribution of our paper can be summarized as follows:

a) We collected two datasets (ASL-10) & ASL-24 by recording sEMG signals generated corresponding to ASL sign gestures and proposed a reliable machine learning pipeline for ASL gesture recognition using an optimal number of sEMG sensors.The pipeline achieved excellent average classification accuracy of 99.99% (ASL-10) and 99.91% (ASL-24), using two and eight SEMG channels, respectively on the two collected dataset.
b) Evaluated around 450 features for each sEMG channels and applied a new ensemble feature selection technique to list the most representative feature subsets to classify ASL hand gestures.
c) Validated the proposed pipeline on the benchmark dataset (Ninapro data-set 5 exercise A) having 12 similar gestures and achieved classification accuracy comparable to state-of-the-art methods.

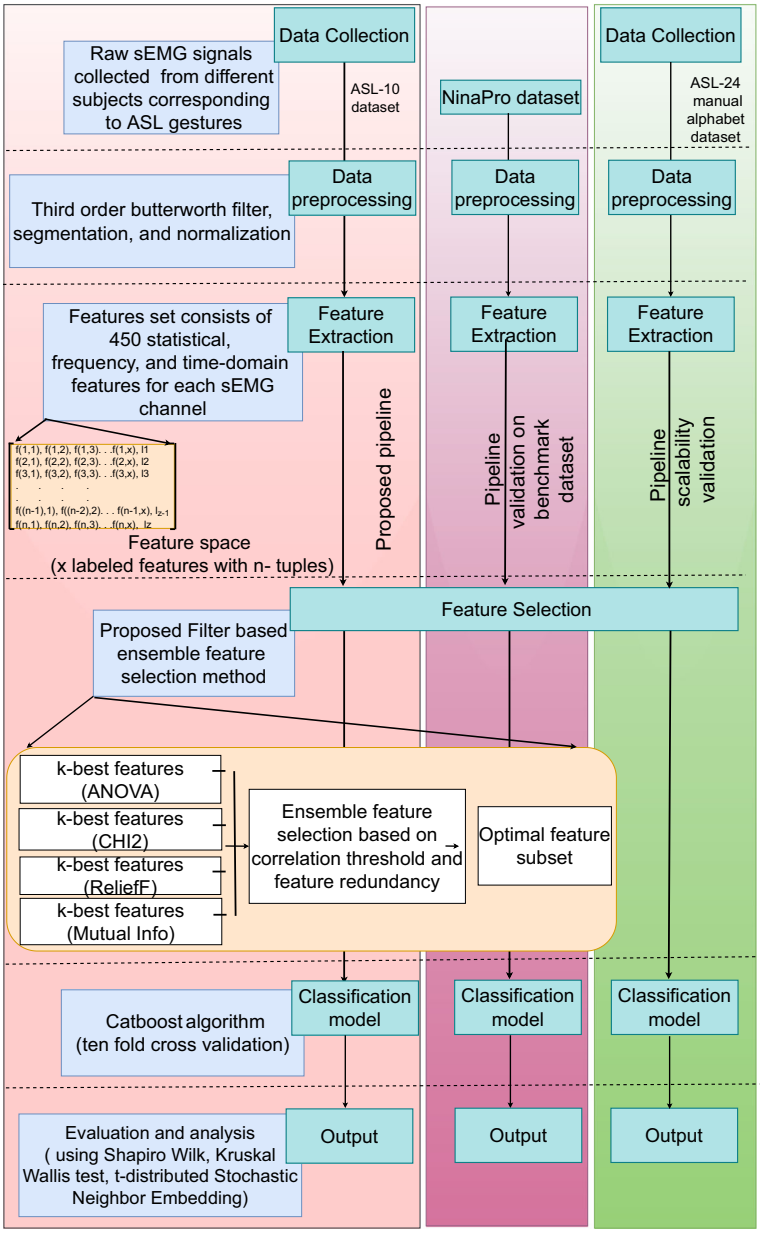

Fig. 1 A schematic diagram showing the methodology used for the proposed machine learning pipeline for ASL recogntion task

The rest of the paper is organized as follows: Section 2 describes related work published from 2004 to 2021. Section 3 describes the internals of the Materials and Methods, dataset description, data preprocessing, and the proposed ensemble feature selection method.

Section 4 lists the experimental results. Whereas in Section 5, we provide a insightful discussion of the results obtained and the limitations of our proposed method. Finally, Section 6 concludes the article with significant findings and future scope.

2 Related work

Development in Visual sign language was started in the 1980s. Visual sign language recognition techniques mainly depend on features derived using hand, face, and body poses. Initial work in this domain applied neural networks along with various image processing techniques for feature extraction in order to recognize the different sign language gestures [64, 105]. Later inspired by the success of deep learning in image processing, researchers exploited Convolution Neural Networks (CNN) to build robust and accurate SLRS [35, 47, 96]. CNN's were very successful while dealing with static image frames but can hardly accommodate the sequence information in continuous image frames [77]. For this purpose, researchers proposed hybrid ML models where they combined CNN with deep learning models such as LSTM and RNN to deal with the sequence information [38, 55, 59, 80, 82]

The efficiency of the visual-based SLRS was enhanced with the advent of depth cameras and accurate motion trackers such as Kinetic sensors and Leap motion sensors [95]. These vision-based gesture sensing devices are efficient in tracking body movements. The main difference between the two is that Kinetic sensors can track the whole body while the Leap Motion sensors are efficient in tracking just the hand gestures [29]. Leap motion sensors use single/multiple cameras and optical motion sensors(usually LED-based), which enable them to capture the gesture information precisely in millimeters. An ASL recognition system based on Leap Motion sensor was developed by Chuan et al. [18]. Using Support Vector Machine, their model achieved a maximum classification accuracy of 79.83%. Another prominent work was proposed by Lee et al. [55], where the authors used Leap Motion sensor and applied a deep learning model (Recurrent neural network) for ASL recognition. They claimed that their model achieved a classification accuracy of 99.44%.

Kinetic sensors cover around 20% of the total research work performed for SLRS development [95]. With the help of their cameras and skeleton stream, devices with kinetic sensors are capable of capturing the motion and depth information of gestures for different sign languages [108]. This information helps to extract even more refined features that contribute in building a more robust, multi-modal SLRS as compared to one obtained after using only vision-based features [13, 34, 91].

Despite wide use, visual and depth-based recognition systems have certain limitations. For instance, their efficiency is dependent upon variables such as camera viewpoint, resolution, surrounding lighting conditions, and occlusion present [10, 52]. It is difficult to control these variables in a real-time scenario. Meanwhile, the Leap Motion controller has varying sampling frequency and needs proper preprocessing in practical scenarios [55]. Another limitation while dealing with these approaches is that a line of sight of the gesture with the camera has to be maintained for the smooth functioning of the SLRS.

The sensory glove-based sign language recognition systems have been a popular choice among sensors-based methods. Various sensors are placed over the hand-worn glove that can monitor physical features associated with sign language gestures. Ahmed et al. [2]. Prominent work using Sensory gloves is presented by Oz et al. [69], where the authors applied sensory glove and motion sensors to develop an ASL recognition system. Using Artificial Neural Network (ANN) and histogram of feature vectors achieves a classification accuracy of 95%. While in a similar work by Oz et al. [70], authors used a noise reduction mechanism in velocity network and extracted local and global features for improving the recognizing

accuracy for ASL gestures. Fels et al. [28] uses a data glove with two optical sensors along with a "Polhemus" sensor to efficiently recognize differnt hand movements . The model was able to achieve an error rate of less than 1.7%. While Mehdi et al. [61] applied gloves with seven sensors for ASL recognition and achieved 88% recognition accuracy. In another work by] Kuroda et al. [54], the authors applied 31 contact and Inductcoder sensors on a custom-built Glove and achieved an accuracy of 85% for different gestures. Oz et al. [68] applied CyberGlove with motion sensor and strain gauge to measure angle for ASL gesture recognition and achieved 96% recognition accuracy. However, the limitation of sensory motion gloves is that they may be difficult to use when deployed with a real-time SLRS application. These gloves cover most of the users palm and fingers, restricting the user from performing other tasks.

Wearable sEMG sensors, that are cheaper than visual and sensory gloves, have evolved as an alternative to develop SLRS. To achieve better recognition accuracy these sensors are often fused with inertial sensors. Some of the prominent work of ASL recognition using sEMG sensors includes the work of Savur et al. [86] where they extracted various time and frequency domains features from the raw sEMG signals. Their experiment used Myo armband with eight sEMG channels and achieved a classification accuracy of 61.04%. In thesis work of Jackson Taylor [92], three different sensors viz. accelerometer, gyroscope, and sEMG sensors (total 15 different channels), were applied for classifying different ASL gestures. Using k-NN and dynamic time windowing, the author achieved an accuracy in the range of 94%-98%. Similarly, Paudyal et al. [71] applied two wrist-worn devices consisting of sEMG and inertial sensors to recognize the ASL signs using Dynamic Time Wrapping, an energy-based approach. By using the combination of all the sensors (34 different channels), Paudyal et al. achieved an accuracy of 97.72%. Wu et al. [101] proposed an ASL recognition system using the sEMG and wrist-worn inertial sensors. They obtained the recognition accuracy of 95.94% Using three different sensors with a total of ten different channels (6 inertial sensor channels with four sEMG channels). An information gain-based filter method was applied for selecting the 30 best features set from 268 features. With the extension of the previous work, the authors applied an auto segmentation technique to identify the occurrence of ASL gesture [99]. Using additional gyroscope channels to the feature vector the proposed method achieved the classification accuracy of 96.16%. In [53], while performing sEMG based ASL gestures recognition, Kosmidou et al. applied Discriminant analysis to find the best features subset from the total number of features extracted. The authors claim to achieve a classification accuracy of 97.7%. In [85], several Time domain features were extracted for the sEMG channels corresponding to different ASL gestures. Using a Support vector machine, their model achieved an offline accuracy of 91.1%. No feature selection method was applied to reduce the feature vector. Their proposed model uses eight sEMG channels, and its performance was evaluated on 2080 samples. Fatmi et al. [27], applied two Myo Armbands and, using their sEMG & inertial sensors (total 26 different channels), proposed another ASL recognition model. In their work, they compared the performance of ANN and SVM with the Hidden Markov Model and claimed to achieve a classification accuracy of 93.79% and 89.05% with these two approaches. With an intent to achieve higher recognition accuracy, most of the sEMG based ASL recognition models applied multiple sEMG sensors; however, increasing the sensors would increase the cost of building an SLRS based on that model. In our proposed work, we have tried to find the optimal number of sEMG sensors for the recognition task without compromising efficiency.

3 Materials and methods

In this section, the proposed machine learning pipeline is systematically introduced along with an introduction to SEMG signals, experimental protocol for data collection, dataset description used for ASL recogntion task .The figure shows the schematic diagram of the proposed method. Initially, a number of data preprocessing measures are taken to deal with data inconsistency, noise or artifacts from the raw sEMG signals.Then appropiate digital filter and data segmentation is performed. A exhaustive feature set containing relevant sEMG features listed in literature are extracted for each window segments. An ensemble feature selection approach combining four diverse filter-based feature selection methods is used to reduce the high dimensional feature vector. Initially, k-best features are extracted from filter-based methods using the feature ranking method. Then, a greedy search approach is used to select subsets from these various filter-based approaches based on the feature ranking and feature importance heuristic. Later, a newly proposed feature combiner that exploits feature–feature and feature–class correlation thresholds is used to combine feature subsets formed across the ensemble. Finally, Catboost algorithm, which perform gradient boosting on decision tree, is trained on the reduced feature subset obtained from ensemble feature selection and validated using 10 fold cross validation.

3.1 Surface electromyography (sEMG)

EMG is a technique to register the electrical activity of muscles during contraction. EMG has the capability to represent human motion activities by generating distinct signal patterns. EMG signal patterns for different hand actions can be easily classified to recognize numerous hand gestures for either prosthesis application or sign-language recognition. EMG can be performed through invasive or non-invasive methods. Invasive technique requires the insertion of needles within the skin surface which is quite painful. sEMG is preferred as it the non-invasive practice which employs electrodes on the skin surface to record the muscle activity. Therefore, in this work sEMG signals for different hand gestures were utilized to classify different digits of ASL.

3.2 Subjects and experimental protocol

The experimental dataset was collected from different subjects(upto 20) of the age of 22-28 years, comprising fifteen males and five females. An ethical approval was taken from the ethical committee, Institute of Medical Science, Banaras Hindu University, Varanasi before taking surface EMG signals from the subjects. A Myo Armband from thalmic lab having an eight-channel wireless sEMG sensor was used to register different gestures for American Sign Language (ASL). Different hand activities for ASL (for ten different digits from 0 to 9) and 24 ASL manual alphabets have been provided in Figs. 2 and 3. The subjects were instructed to wear the armband in the lower forearm near the elbow region. The muscles (i.e. flexor, extensor as well asbrachiradialis) towards the lower forearm are directly responsible for the movement of wrist and fingers. Subject's hair were removed from the target portion of the forearm and also the skin surface was cleaned with alcohol. During placement of the sensors, efforts were made to maintain a fixed distance from the elbow of different subjects. The experimental setup for acquiring EMG data is described in Fig. 4

Fig. 2 Different hand gestures for digits 0 to 9 in ASL

3.3 Data acquisition and sensor placement

For each ASL gesture, 5 sec continuous EMG data were collected from all the participants. All the data were acquired at a sampling frequency of 200 Hz. Myo data capture software interface was employed to acquire the EMG data on the PC and Myo web based interface was used to monitor the sEMG signals. To minimize the effect of muscle fatigue during data collection a fixed protocol depicted in Fig. 5 was followed. The data collection was done in

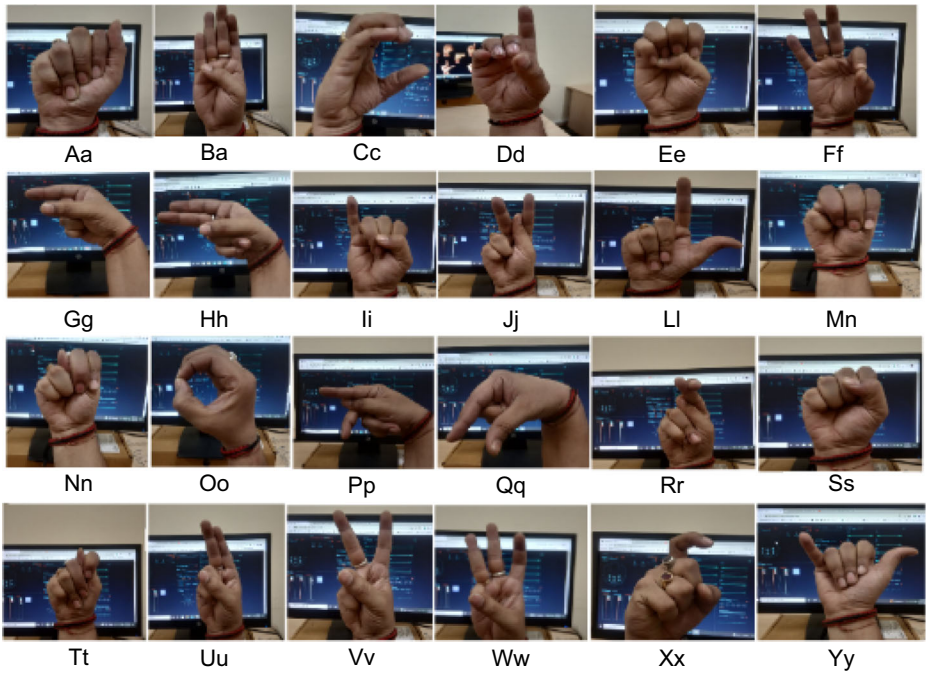

Fig. 3 Figure illustrating the different ASL manual alphabets used in ASL-24 dataset

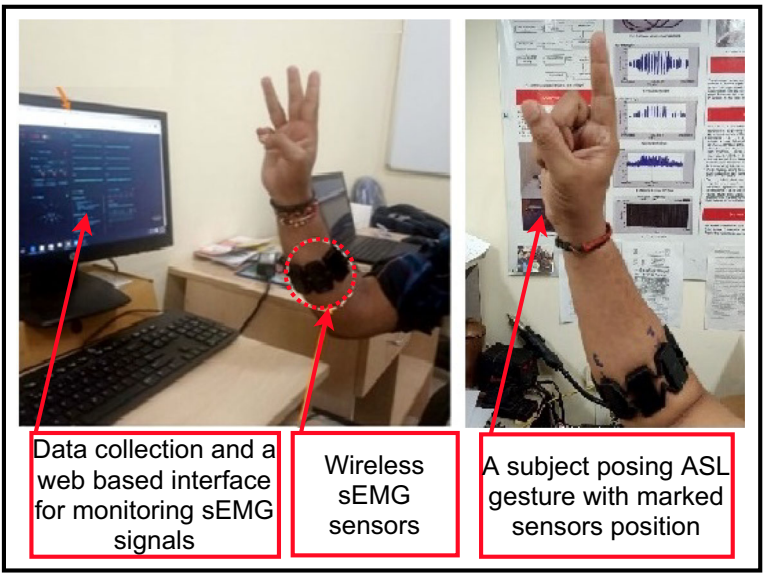

Fig. 4 Data collection setup

various trials and sessions. In each session, different trials of 5 sec were performed for each ASL sign. Moreover, an appropriate gap was maintained between successive sessions.

EMG signals are dependent upon the placement of the sensors used for data acquisition. Initially, the all the Eight sensors of the Myo Armband were provided with identification marks from 1-8 and effort was made to place the sensors marked with 1 (with blue light) on Extensor carpi ulnaris muscles. Further, care was taken to put an sEMG sensor on Extensor carpi radialis longus by adjusting the armband length. These two muscles were initially identified and then marked. The sensors corresponding to these specific muscles were identified and accordingly the respective sEMG channels were annotated in the dataset. The placement of the two specific sensors was inspired from the work performed by Pizzolato et al. [100].

3.3.1 ASL digit dataset (ASL-10)

The dataset consist of sEMG signals collected from 20 healthy subjects of age 22-28 years, comprising fifteen males and five females. Each of the subjects were made aware of ASL digits gestures through videos and proper illustrations. Using the data acquisition and sensor placement protocol, each of the subjects are made to generate gesture for each of the ten ASL digits. For ASL-10 dataset, the data was collected and processed for only two sEMG sensors. Though the sampling frequency for the device was lower, still total 53000 samples were collected and analyzed. For, two sEMG signals around 900 features were extracted. The raw sEMG signal pattern for a subject posing hand gesture digit 9 of ASL has been shown in (Fig. 6).

3.3.2 Ninapro reference dataset

The Ninapro DB5 dataset has been used as a reference dataset to evaluate the performance of classification algorithms applied over sEMG signals [6, 73]. This dataset consists of sEMG signals collected from 10 healthy subjects over 53 different movements including a neutral hand position. These 53 different movements were divided into three different exercises, collected using Myo Armband. We compared our results with the exercise A of this dataset that consists of 12 different gestures including finger movements.

The dataset is collected by using two Myo Armbands containing 16 sEMG channels. However, our proposed pipeline was analyzed using 2 sEMG channels placed at specific muscles. To compare our results with the reference dataset, we selected the two most significant sEMG channels from the Ninapro dataset. For this selection, we applied our proposed feature selection approach on 16 channels and chose the two most significant sEMG channels.

3.3.3 ASL manual alphabet dataset (ASL-24)

To analyze the scalability of our machine learning pipeline, we created a separate dataset containing an increased number of gestures (24 ASL manual alphabet). New ASL signers use these ASL manual alphabets in fingerspellings words, where each alphabet of the word is generated by making different hand shapes. There are twenty-six ASL manual alphabets corresponding to 26 letters (A-Z). Most of these ASL alphabets are made using a single hand. However, out of 26, We have only considered 24 letters (static signs of ASL), excluding letters 'j' and 'z,' as these two letters involve dynamic hand movement during fingerspelling generation. Figure 3 illustrates the 24 ASL manual alphabet used for the datatset

The dataset was collected from 10 subjects consisting of eight men and 2 female of age (22-28 years). Effort was made to keep the experimental protocol, data acquisition, and dataset preparation steps similar to the ASL-10 data-set. However, considering the increased number of gesture we deployed eight sEMG sensors for collecting sEMG signals corresponding to 24 ASL manual alphabet gestures. These eight sensors were placed on the right hand of each subjects such that two of the sensors are placed at Extensor Carpi Ulnaris and Extensor Carpi Radialis Longus muscles. Total 83000 samples for 24 ASL gestures were collected and processed. For eight sEMG signals around 3600 features were extracted.

3.4 Data set preparation

For the experiment, two new datasets consisting of raw sEMG signals corresponding to ASL gestures was collected. Signals corresponding to various sEMG sensors were collected simultaneously and stored in CSV file format. These sEMG signals can be represented as a multivariate time series, with each of the sEMG channel acting as a separate variable. Suppose $T_s(u)$=$\{u_1, u_2, u_3, u_4 ... u_m\}$ be a time series corresponding to each sEMG sensor with m data points, s be the total number of sEMG sensors, $W_s(p, s)$ be the window segment of length p (p data-points of s sEMG channel), z be the total number of samples, and l_i be the annotation; $0 \leq l_i \leq 9$. Raw sEMG signals collected can be depicted as $D_{raw} = \{(T_s(u_1), T_s(u_2), T_s(u_3) \cdots T_s(u_s), l_i\}$; $T_s(u) \in \mathbb{R}^{1*n}, l_i \in \mathbb{R}$. For each window size $W_s(p, q)$, various relevant features were calculated and stored. For classification, we framed 10 and 24 classes (C_i), one corresponding to each gesture for both the datasets

(ASL-10 and ASL-24). Classes C_i can be defined as:

$$C_i = \left\{ f_{(j,1)}, f_{(j,2)}, f_{(j,3)}, f_{(j,4)} \cdots f_{(j,h)}, l_i \right\};\ C_i \in \mathbb{R}^{N*(h+1)}, 1 \leq j \leq N$$

Where, $f_{j,k}$ is feature extracted for window segment $W_s(p, q)$, h being total number of features, C_i represents different classes each mapped with a unique ASL gesture; $1 \leq C_i \leq Z.Z$ being 10 and 24 for ASL-10 and ASL-24 datasets, respectively.

3.5 Data preprocessing

The different sensors present in the armband were responsible for capturing separate EMG signals from distinct muscle locations of the lower forearm. Raw sEMG signals from all the channels were recorded in CSV format where each column corresponds to an individual channel. Initially, the basic data processing techniques such as identifying & handling missing values were applied on the raw signal collected for the ASL gestures. After feature extraction, the features values were transformed in the range of 0 - 1 using Normalization.

3.5.1 Filters

To deal with the noise that occurred during data acquisition, a digital third-order Butterworth bandpass filter of bandwidth 5-500 Hz was applied on the sEMG signals. Moreover, as suggested in article [85], a configurable Butterworth filter was used to eliminate 50Hz power line noise.

3.6 Data segmentation

Data segmentation is used to divide the sEMG data stream into fixed sizes, also referred to as window size, to evaluate the features over the data points contained in windows. This window acts as a unit size for which the features are collected in time series. Choosing the size of window plays a vital role as considering a smaller window may not represent a complete gesture and a larger window size may lead to computational overhead and unnecessary delay in real time applications.

To find the optimum window size, classification was performed on various lengths of data segments. We applied classification algorithm on 400, 300, 200 and 60 number of data samples per window size. Separate analyses was done using the same window size with a 50 percent overlap. Considering the tradeoff between computational delay and accuracy, we found an optimal solution at window size of 60. Data was collected on 200 Hz sampling rate and keeping the window size 60, enable us to maintain delay comparable to permissible limit for real-time gesture recognition. Hudgins et al. [42]

3.7 Feature extraction

For the sEMG channels, large number of features from the frequency, time, and frequency-time domain were extracted to accommodate the relevant features used in biomedical signal analysis [50]. Statistical features such as Kurtosis, Mean, Variance, Absolute energy, Autocorrelation, and Standard derivation were also calculated. Time domain sEMG signals were transformed into the frequency domain using Fast Fourier Transform (FFT). Because various ASL gestures would produce different frequency distributions, we chose 95 FFT coefficients from 5 to 100 as features [49, 86, 106]. Apart from these other features such

as Entropy, Benford Correlation, c3 non linearity statistics [87], Complexity-invariant distance (complexity information) [8], Mexican hat wavelet [93], Spectral centroid (absolute), skewness, and the kurtosis of the Fourier transform, Power spectral density based on Welch Method [45], Lempel-Ziv based complexity [1], One dimensional Matrix profile [102], Partial auto correlation, Root Mean square, Sample entropy, Friedrich coefficients [31], Absolute sum of changes (consecutive time series value changes, Langevin fixed point (Largest point of deterministic dynamics) and Quantiles were extracted for each of the sEMG signals. A resultant feature set was formed by combining the features of all sEMG channels.

3.8 Features selection: using proposed ensemble method

We extracted a sufficiently large amount of features from the raw signals collected to find the best representative feature subset for classifying the ASL gesture registered using sEMG signals. However, the extracted features resulted in a higher dimensional feature vector for the classification task. So to avoid the effect of the curse of dimensionality [44] and redundancy, a new ensemble feature selection technique was applied to find the most representative feature subset from the larger feature space. Filter-based feature selection methods such as ANOVA, Chi-Square, Mutual Info, and ReliefF were chosen for the ensemble approach. These methods are independent of the use of machine learning algorithms and computationally efficient compared to wrapper-based methods [12, 39].

Ensemble feature selection has been applied to various problems in the field of network traffic analysis, Biomedical signal processing, pattern recognition, etc., involving the optimization of larger feature spaces [15, 81, 84]. The use of ensemble feature selection can be justified by the fact that the applying single feature selection method may result in providing a local optimal feature subset or can have baisness on selected feature subset. Hoque et al. [41]. Integrating diverse feature selection models to form ensemble models helps to overcome individual feature selection model biases and achieve improved performance while classifying a wide range of applications .However, arbitrarily increasing the number of individual feature section model for creating ensemble may not necessarily produce improved results. One such observation was made by Wang et al. [97], in their work they performed an empirical study consisting 17 different feature ensembles made form combining multiple feature ranking techniques. The authors claims that ensembles of few rankers perform comparable or better than ensemble made form multiple or higher number of feature ranking techniques.

Olsson and Oard [67] applied the ensemble selection method on text classification and achieved improved Precision and F1. A similar observation was reported by Yu et al. [103], in which the author reported an increase in performance by using the Genetic algorithm-based ensemble method.

Zang et al. [107] proposed a cost-sensitive feature selection model which applied multi-objective particle swarm optimization for maximizing classification performance using minimum costs associated with features. Miften et al. [62] proposed an ensemble feature selection technique to identify the most influential features for classifying six different EMG signal-based hand-grasp. The authors combined three popular feature selection methods (Chi-squared, Mutual info, and RFE) to create an ensemble and uses a ranking combination approach to obtain an integrated feature set. Further, the Fisher discriminant ratio was applied to determine the threshold value, which eventually helped attain the most significant features subset for the classification task. The proposed ensemble method achieved a classification rate of 98.5% (average for five subjects) and exhibits better results than

many previous research works. In addition, the authors claimed to achieve about 7% higher classification result than using the classification model with the single feature selection model.

Our proposed work of ensemble feature selection is inspired by the ensemble technique used in Miften et al. [62] described earlier in this section. The authors applied a ranking-based combinatorial approach to the overall feature set of three different feature selection techniques (Chi-Square, Mutual Info, and RFE) used in their work. However, feature ranking methods are not considered efficient in managing redundant variable [14, 33, 40].

We proposed a new approach to combine the feature set of different individual feature selection methods for our ensemble feature selection method. The proposed combiner aggregates the feature sets to obtain an optimal feature set by incorporating feature-feature and feature-class correlations property. Feature-class correlation is defined as the relevance of a feature based on the correlation between the feature and the ASL class label. The two concept feature-feature and feature-class correlation helps to achieve relevant and non redundant features form the large feature vector obtained in feature extraction step.

The proposed ensemble selection can be explained in two major steps. In the first the filter based methods(ANOVA, Chi-square, Mutual Info, ReliefF) provides individual k-best feature set.The choice of using four filter method for creating ensemble was based on findings of Wang et al. [97]. while, in the second step of ensemble, these feature set are aggregated using the feature-feature and feature-class relation providing the optimal feature subset.

In our work an optimal feature set, $final\_selected$, is initially made by selecting distinct features from the four filter based feature selection methods(ANOVA, CHI2, Mutual Info, ReliefF). In this step, using the greedy search approach, the higher ranking features common in all four feature set are directly added to the optimal features set, $final\_selected$. The selection of common features from all four feature set is based on our heuristics that these features are the highly significant feature with efficient discriminatory properties.

Later, the combination of any three individual feature selection method out of the four methods are listed as groups. The non redundant and common features in the the listed group are identified.The identified features are only added to the optimal set $final\_selected$, if they satisfied the correlation threshold.

In the next step, the distinct features selected in at any two of the four feature selection methods is added to the optimal feature set after validating through feature-class threshold. A correlation test is performed on the optimal feature set, and the features having correlation above a threshold ($thres_1$) are truncated off from this feature set.

All those features not initially selected in the $final\_selected$ are combined and optimized based on the correlation threshold($thres_2$) and stored as $rem\_features$. Further, for each feature f_i in $rem\_features$; the feature-feature and feature-class correlation with the features in optimal feature set $final\_selected$ and ASL classes is performed. If the feature f_i satisfies thresholds, it is added to the optimal feature set. The thresholds $thres_1$, $thres_2$ and $thres_3$ are derived based on the empirical results carried out with aim to attain higher classification accuracy. We used the threshold $thres_1$and $thres_2$ equal to 0.85 while the $thres_3$ as 0.90 based on the exhaustive experimental result. While the initial number of "k" features is selected based on the heuristics and the empirical analysis. The ensemble feature selection method is illustrated as Algorithm 1. The proposed feature selection approach results in feature subset consisting of best discriminatory features form each sEMG channel.

Input: Three features sets consisting of k features each for ANOVA, Mutual Info, ReliefF and Chi Square:
$FS_{chi2} = \{f_1, f_2, f_3 ... f_k\}$; ▷ k-features set selected using Chi-square
$FS_{mi} = \{f_1', f_2', f_3' ... f_k'\}$; ▷ k-features set selected using Mutual Info
$FS_{anova} = \{f_1'', f_2'' ... f_k''\}$; ▷ k-features set selected using Anova
$FS_{ReliefF} = \{f_1''', f_2''' ... f_k'''\}$; ▷ k-features set selected using ReliefF
$thres_1, thres_2$ ▷ Feature-Features correlation thresholds
$thres_3$ ▷ Feature-Class correlation threshold

Output: Final selected features subset

Initialization
feature_selection_methods = $[FS_{chi2}, FS_{anova}, FS_{ReliefF}, FS_{mi}]$
$final\_selected = \phi$ ▷ Selected features set initialized as empty
$rem\_features = \phi$ ▷ Features set initialized as empty
$final\_selected = FS_{CHI} \cap FS_{MI} \cap FS_{ANOVA} \cap FS_{ReliefF}$ ▷ Common features in all the feature selection methods
function FIND_FEATURES(X,Y)
    features = (Set X ∩ Set Y - ($FS_{chi2} \cap FS_{mi} \cap FS_{anova} \cap FS_{ReliefF}$))
    return features
end function
function SEL_UNCORR_FEATURES(feature set,threshold)
    // compute correlation matrix of (feature set) and drop the features from the (feature set) based on user defined threshold
    return features set
end function
for each distinct pair (X,Y: X ≠ Y) in feature_selection_methods list **do**
    $final\_selected = final\_selected \cup$ FIND_FEATURES(X,Y)
end for
$final\_selected$ = SEL_UNCORR_FEATURES($final\_selected, thres_1$)
$rem\_features = (FS_{chi2} \cup FS_{mi} \cup FS_{anova} \cup FS_{ReliefF})$ - $selected\_features$
$rem\_features$ = SEL_UNCORR_FEATURES(rem_features, $thres_2$)
for each features f_i in $rem\_features$ **do**
    //evaluate correlation coefficient C_i with all the features in $selected\_features$
    //evaluate correlation coefficient D_i with Classes
    if ($C_i \leq thres_1$ and $D_i \geq thres_3$) **then**
        add f_i in $selected\_features$
    end if
end for
$selected\_features$ ▷ Final selected features subset

Algorithm 1 Ensemble feature selection.

3.9 Classification algorithms

The Catboost algorithm is used to classify the different ASL gestures associated with three data sets (ASL-10, ASL-24, and Ninapro reference dataset). Catboost is a machine learning algorithm that performs gradient boosting on symmetric decision trees. Gradient boosting

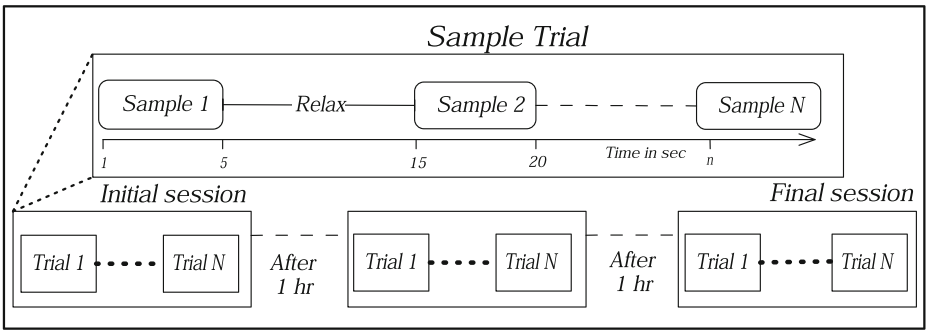

Fig. 5 Data collection protocol

is based on the concept of how strong predictors can be produced by iteratively combining weaker predictors through greedily applying gradient descent. Moreover, the use of the symmetric property helps the algorithm to deal with the model over-fitting, improves the prediction time and to be more reliable to parameter change [24, 75]. The catboost algortihm works well with heterogeneous data and have achived many state of art classification accuarcy on various benchmark dataset [24].

3.10 Evaluation metrics

We evaluated the performance of the proposed machine learning pipeline with a focus on measuring its applicability if used as a recognition module in a sign language recognition system. The efficiency of such a recognition module mainly depends upon factors such as recognition time, recognition accuracy, pervasiveness, scalability, invasiveness, etc. [71]. However, in this article, we framed the experiments focusing on evaluating recognition time, recognition accuracy, and system scalability.

For measuring the recognition accuracy we framed the ASL gestures recognition task as a multi-class classification problem. Along with accuracy we evaluated other widely accepted metrics for multi-class classification [36]. Metrics such as Receiver Operating Characteristic (ROC) curve, Kappa score ,Matthews correlation coefficientMCC_{mc} were used to measure the performance of the experiments. Most of these evaluation metrics are derived based on true positive(tp), true negative(tn), false positive(fp) and false negative(fn) samples and are defined as follows:

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{1}$$

Accuracy is a measure of how accurately the classification model is making predictions on a given dataset.

MCC is a reliable statistical metric to measure the classification results. The MCC score archives better results if the classifier performs well with all the true positive, true negative, false positive, and false-negative samples and effectively handles imbalanced datasets [16]. It is defined as:

$$MCC_{bi} = \frac{tp_{(bi)} * tn_{(bi)} - fp_{(bi)} * fn_{(bi)}}{\sqrt{(tp_{(bi)} + fp_{(bi)})(tp_{(bi)} + fn_{(bi)})(tn_{(bi)} + fp_{(bi)})(tn_{(bi)} + fn_{(bi)})}} \tag{2}$$

Where, $tp_{(bi)}$, $tn_{(bi)}$, $fn_{(bi)}$, $fp_{(bi)}$ refers to true positive, true negative, false positive and false negative when framed as a binary classification problem. However, this concept

can be extended for multi-class classification. The multi-class MCC_{mc} can be derived using multi-class confusion Matrix, C, and is given as:

$$MCC_{mc} = \frac{f * e - \sum_n^N p_n * t_n}{\sqrt{(e^2 - \sum_n^N p_n^2) * (e^2 - \sum_n^N t_n^2)}} \tag{3}$$

Where, C is the confusion matrix of size (n∗n), $t_n = \sum_p^N C_{pn}$ denotes the number of times class 'n' correctly occurred, $w_n = \sum_p^N C_{pn}$ is the number of times class 'n' was predicted, $f = \sum_n^N C_{nn}$ denotes the total number of instances correctly predicted. While, $e = \sum_p^N \sum_q^N C_{pq}$ denotes total number of samples.

Another metric used to evaluate the performance was Kappa [30]. The metric help to measure the agreement between real and assigned classes by a classifier. It can also be defined using multi-class confusion Matrix, C, similar to MCC_{mc} and is given as:

$$Kappa_{mc} = \frac{f * e - \sum_n^N p_n * t_n}{e^2 - \sum_n^N p_n * t_n} \tag{4}$$

Receiver Operating Characteristic (ROC) curve known as AUC is also used as a metric for performance evaluation of proposed pipeline. AUC represents a modal ability to differentiate between classes.

Further to estimate the recognition time of the module, we calculated the time spend by trained classifier to predict a single ASL gesture and termed as Response Time(T_{Res}). Let $W_s(s, p)$ be the optimal window size used in the model. The Response time is denoted as

$$T_{Res} = T_{w_s(s,p)} + T_{fe} + T_{Recog}$$

T_{w_s} is the time spend to extract a window segment $W_s(p, s)$ from raw input signal. T_{fs} is the time spent on extraction of selected features from the input data window segment $W_s(p, s)$. T_{Recog} is the recognition time, which is the time spend to predict the label for the window segment $W_s(p, s)$.

Finally, to measure the scalability of the proposed pipeline, we framed a new dataset, ASL-24, with an increased number of gestures and evaluated its performance on this dataset to validate its scalability.

4 Results

The proposed experiments were performed with the objective to attain higher classification accuracy for sEMG based gesture recognition task. During the experiments we utilized mainly three datasets ASL-10 (10 ASL digit gestures), NinaPro reference dataset(12 hand gestures), ASL-24(24 ASL manual alphabet gestures).The proposed pipeline is evaluated on these dataset with different objectives. ASL-10 is used to measure the efficiency of our proposed pipeline using optimal number of sEMG sensors for 10 ASL sign (0-9 digit). While NinaPro reference dataset is used to study the generalizability of proposed method on a benchmark dataset. whereas, ASL-24 dataset is used to validate its scalability.

To evaluate the efficiency of the proposed pipeline on ASL-10 dataset, we ran the model multiple times with ten-fold cross-validation. The pipeline achieved an overall average classification accuracy of 99.99% using Catboost algorithm. In addition, three other popular boosting algorithms, (Gradient Boost, Extreme gradient boost, Light Gradient Boosting),

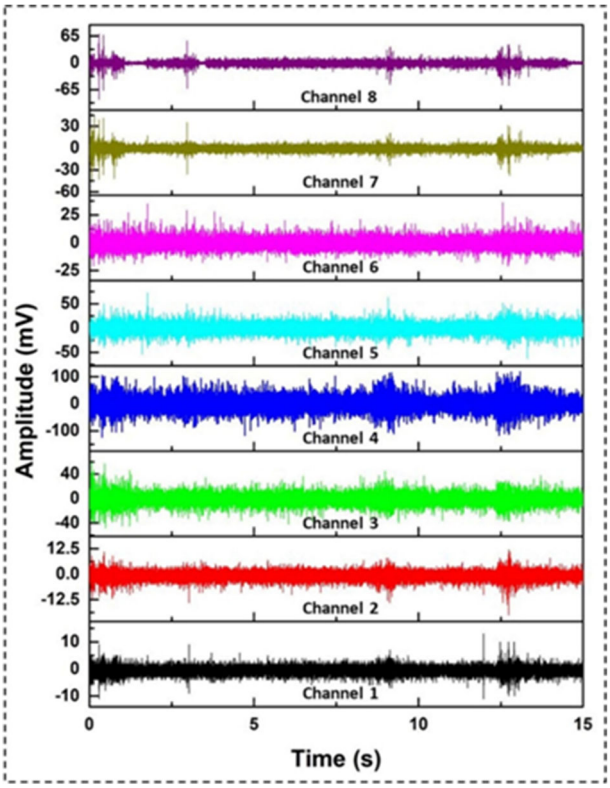

Fig. 6 Raw sEMG signals patterns of a subject for digit 9 of ASL

were applied and evaluated. The majority of the boosting algorithm used produced similar classification accuracy.

In addition, standard evaluation metrics, as discussed in Section 3.10, were also calculated to comprehensively validate the obtained classification results and the efficiency of the pipeline. Figure 7 depicts the multi-class "MCC" score against the number of iterations of the Catboost algorithm for its 10 fold cross validation. In less than 50 iterations, the model achieves the MCC scores greater than 0.9 and, after 150 iterations, eventually achieves the highest MCC score. MCC measures the correlation between the predicted value and true value of instances by the classifier. The high MCC scores suggest that the model performs well with the positive and negative samples [46]. Meanwhile, Fig. 8 illustrates the Kappa values obtained for various cross-validation step, supporting the higher efficiency of the pipeline. The higher Kappa scores suggest the better agreement with the predicted value and true value of instances by the classifier, hence its robustness. In addition, the multi-class AUC [51] for all the ASL classes were plotted and shown in Fig. 9. For our model, the Multi-class AUC reaches the value of 1 in less than 50 iterations. The higher AUC values suggest the ability to distinguish between positive and negative classes.

The "Multi-class" loss function was plotted for each of ten-folds to investigate the model performance. Figure 10 illustrates the "Multi-class" loss function with respect to the number of iterations of the classification algorithm used. The figure illustrates minimal error

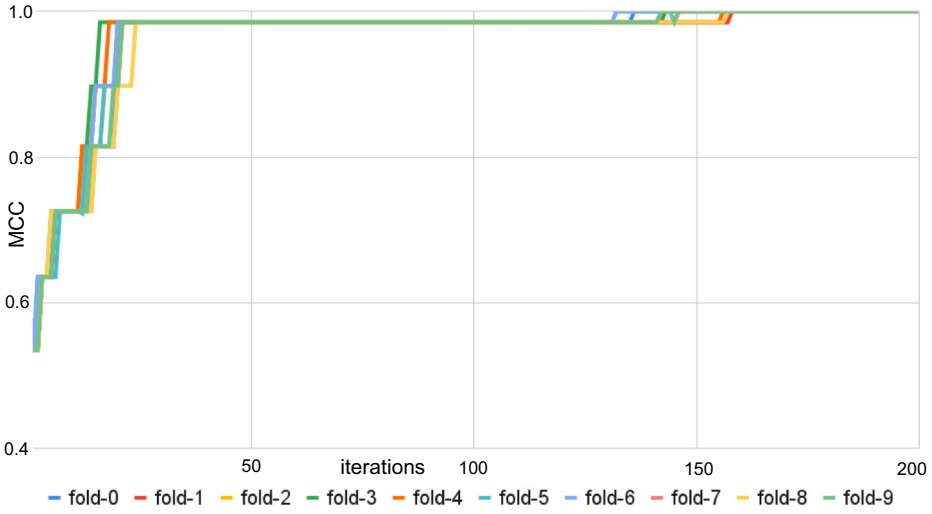

Fig. 7 MCC score plotted against number of Catboost iterations for each of 10 fold cross validation for ASL-10 dataset

produced by the model while predicting the class labels. The "Multi-class" loss function approaches zero after 150 iterations of the Catboost algorithm.

The same pipeline was applied it on a benchmark dataset, Ninapro database, which is used for accessing the sEMG based classification methods. Database 5 of Ninapro consists total of 53 various hand movements and gesture grouped into three different exercises set. For better comparison and to relate the benchmark dataset with our problem, we used only the raw sEMG signals corresponding to 12 different hand gestures of exercise A, of the

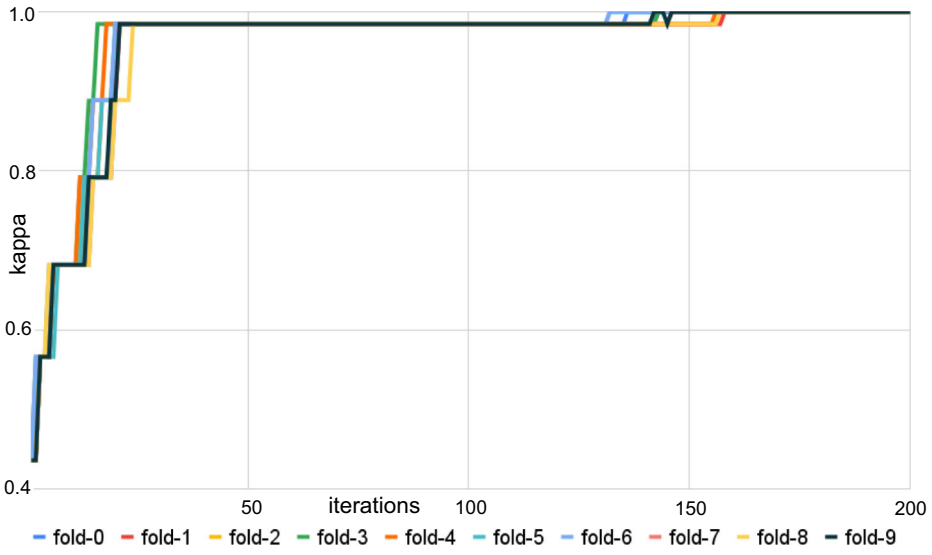

Fig. 8 The Kappa values calculated for multi-classification performed at the ASL-10 dataset

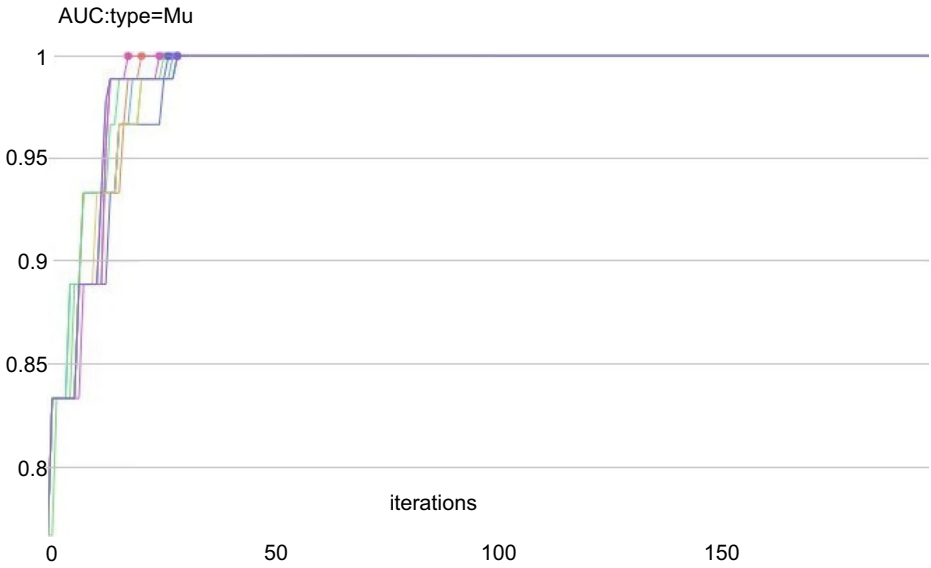

Fig. 9 Multi Class AUC calculated for ten fold cross validation calculated for ASL-10 dataset

Ninapro database 5. The pipeline performed extremely well and achieves the MCC score almost to 1 while classifying the sEMG signals for the Ninapro database. Figure 11 illustrates the accuracy achieved for each of the ten fold validation. As compared to the ASL dataset (Fig. 12) the highest accuracy was achieved in lower number of iterations performed

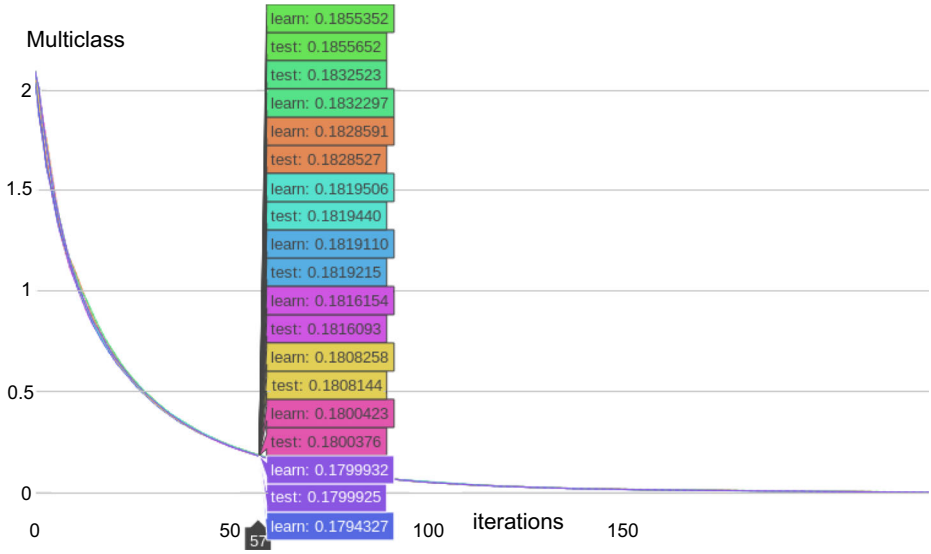

Fig. 10 Figure illustrating the Mutlticlass loss function using 10 fold cross validation(ASL-10 dataset)

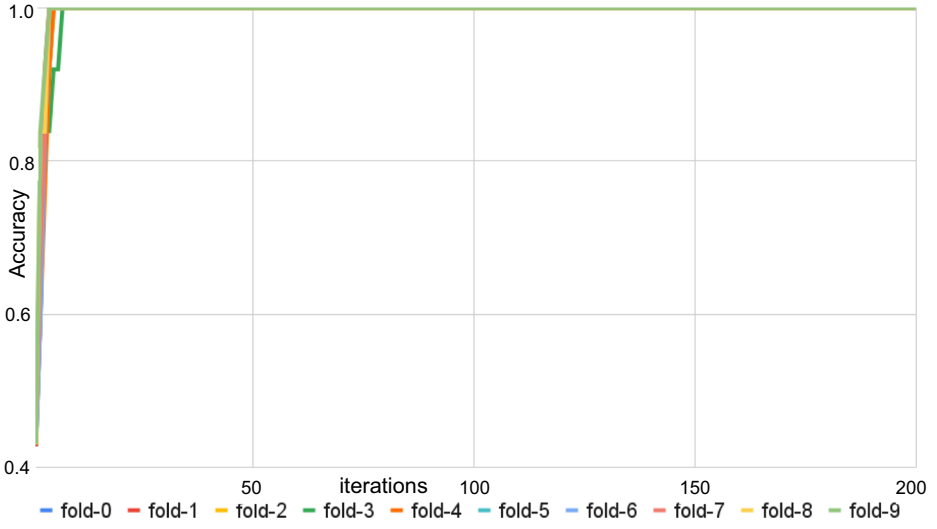

Fig. 11 The accuracy obtained for the Ninapro subset using the proposed pipeline

by the classification algorithm. Normalized confusion matrix describing the pipeline performance in shown in the Fig. 13. For most of the classes, class labels were predicted accurately with negligible miss-classification.

t-distribution Stochastic Neighbor Embedding (t-SNE) [94], was used to map the selected features subset in lower-dimensional for better visualization . The t-SNE is a nonlinear dimensionality reduction technique that efficiently captures the local structure(pattern) in the original space and represents that pattern in the lower dimension [98]. It ensures that if two points are close to each other in higher dimensional space, those points should also

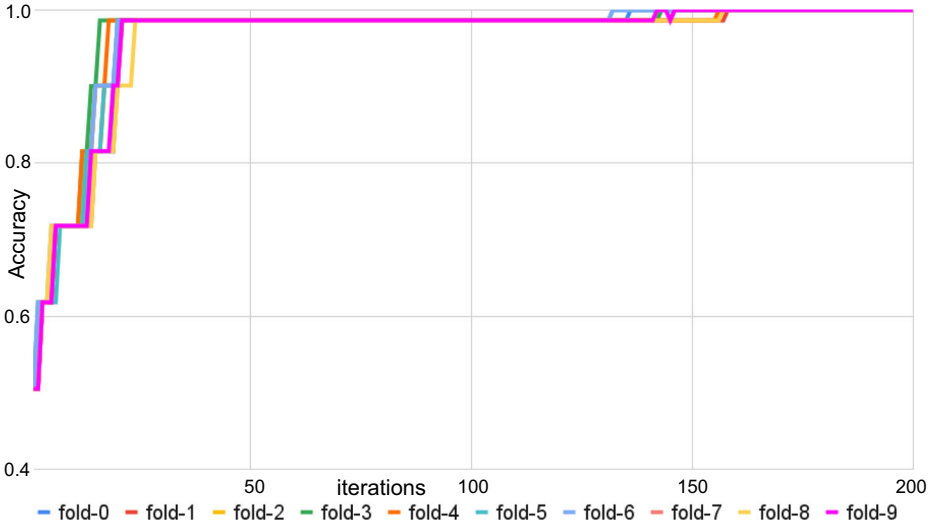

Fig. 12 Figure illustrating the accuracy achieved for ten fold cross validation with respect of number of iterations of classification algorithm

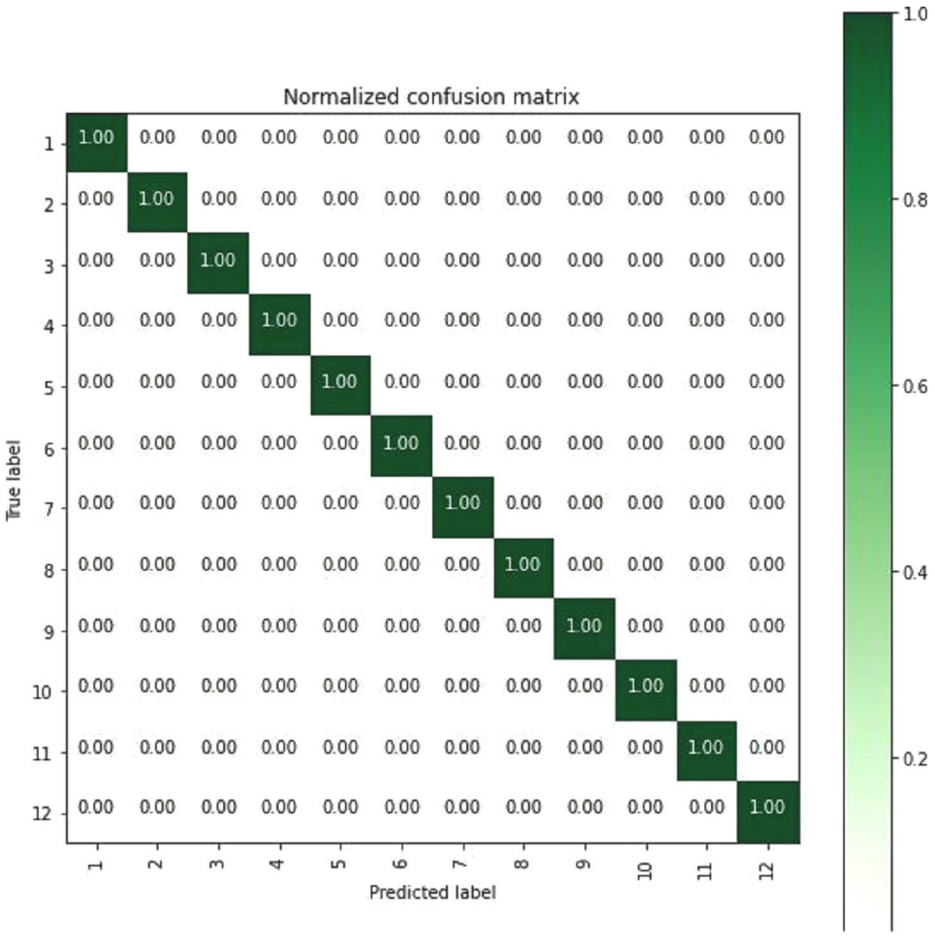

Fig. 13 Normalized confusion matrix for the reference Ninapro Dataset

remain close to each other in reduced dimension. The Fig. 14 illustrates the labeled selected features set for the ASL dataset when visualized through the t-SNE. For the t-SNE plot, values of two hyperparameters variables, "max-iterartion" and "preplexity" were taken as 1000 and 50, respectively. Other Fig. 15 illustrates the t-SNE plot for the features subset obtained for the Ninapro Data subset. Hyper parameters values of "max-iteration" and "perplexity" were made similar to ASL dataset for plotting this graph. The selected features for the Ninapro dataset also depicted identical properties like the ASL dataset forming the almost non overlapping clusters for each gesture class.

Among the features extracted from the raw sEMG signals, the majority of FTT coefficients were listed as the most significant features by the ensemble feature selection. The optimal feature subset, as illustrated in Table 3, consists of various components of FTT coefficients in the range of coefficient 30 to coefficient 36. In another experiment, by increasing the number of features to top 250 in the first level of the ensemble feature selection, we found that the range of the various FFT coefficients in the selected feature set increases from

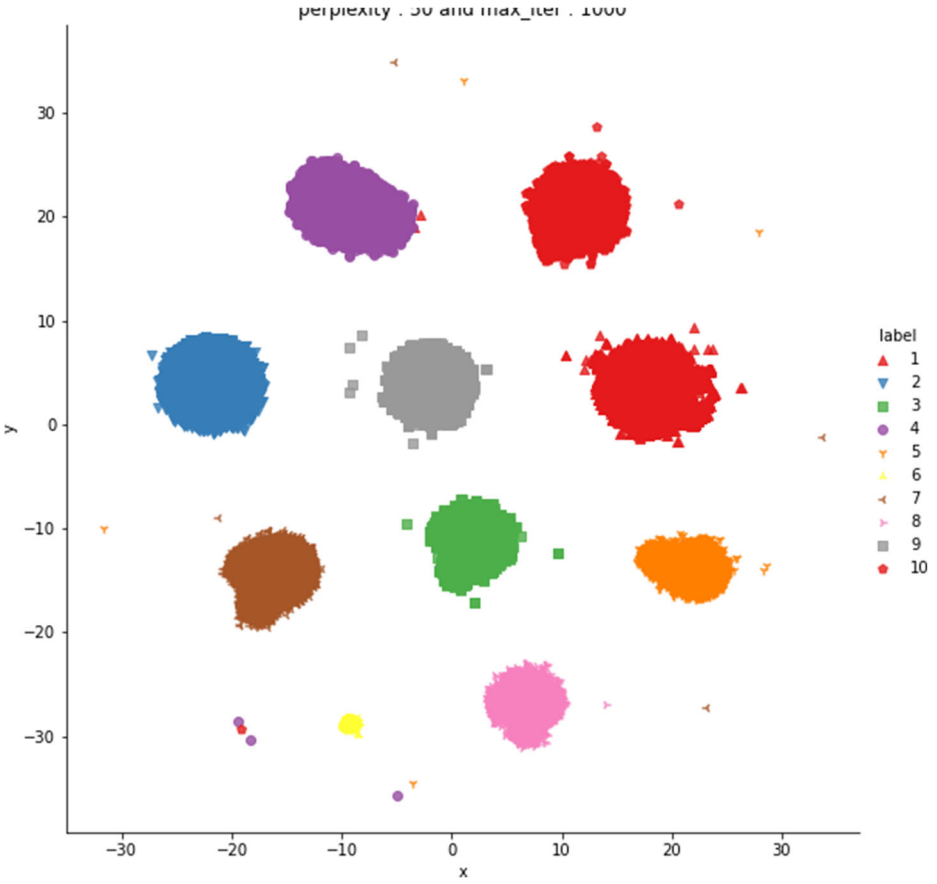

Fig. 14 Features distribution corresponding to ten classes of ASL-10 dataset using t-sne

coefficients 30-36 to coefficients 30-51. The importance of individual features was analyzed in separate experiments. For analyzing the efficiency of each of the FFT components, only single components of the FTT coefficients (REAL, IMG, ABS, ANG), were used in the selected feature set while applying the pipeline. The Table 1 highlights the average classification accuracy achieved through various features. The FFT coefficients combined performed better than the individual components.

To identify the impact of various features on the classification task, we calculated the SHAP values [58]. SHapley Additive explanations(SHAP values) is a technique to explain the factors contributing to predictions made by a Machine learning model. It includes calculation of Shapley values which is based on Game Theory. The Fig. 16 highlights the global importance of the selected features while classifying different instances of the ASL Classes. In the plot, features with larger Shapley values (absolute) are arranged in decreasing order. The larger the Shapley values the larger is the importance of the features for predicting the class labels. The global importance are obtained by averaging the Shapley values(absolute) per features in the dataset.The FFT coefficients 33 (IMG) have the highest impact on all the 10 classes of ASL digits. Similarly, the FFT coefficients 33(ANG) and 35(ANG) are the next influential features having the most significant impact on all the class prediction.

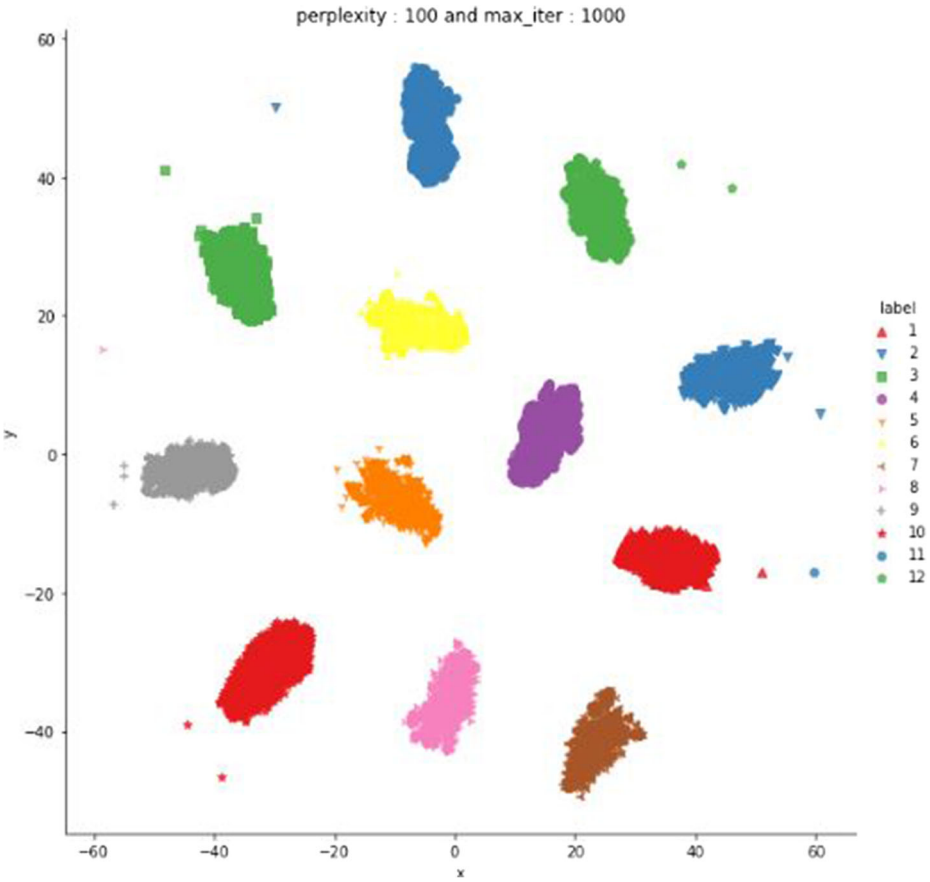

Fig. 15 2D-classwise features projection using t-sne (Ninapro dataset)

In addition, the Fig. 17 illustrates the feature importance and the effect of features value on individual class label prediction. The y axis of the plot are shows the various features involved while classifying the class 1 of the ASL dataset, while the values at the x-axis represent the Shapley values obtained from that features. Meanwhile the colors in the figure reflects the features value from high to low. For the FTT coefficient 35(ANG), the higher is the value the higher is the impact on the class label prediction. Similarly, the effect of other individual feature value on the class prediction can be derived using the figure.

Maintaining a optimal window size is crucial as larger widow size can increase the recognition time in real time application. To find the optimal window size, efficiency of the the proposed classification pipeline was analyzed for different window sizes. The window size of 400, 300, 200, 60 data points with 50 percentage overlap and without overlap was analyzed. However, We achieved the optimal window size of 60 with accuracy 99.99%.

Another experiment analysis was made to measure the Response time of our classification model, which is depicted as the total time incurred from obtaining test data input to predicting its class label (described in Section 2.11). Response time is approximated as the summation of various time delays incurred during window segmentation, feature extraction, and recognition of a single instance. The total Response time turns out to be 316ms.

Table 1 Various features and the corresponding performance metrics

Features	Average accuarcy	Features	Average accuarcy
FTT(ABS)	99.71%	FFT combined(8 channels)	99.99%
FFT (IMG)	99.53%	Other features excluding FFT (8 channels)	97.22
FFT(ANG)	99.49%	Time domain features (8 channels)	51.09
FFT (REAL)	99.42%	Friedrich coefficients (8 channels)	83.33
FFT combined	99.99%		
Time domain features (2 channels)	43.01%		
Other features excluding FFT(2 channels)	96.7%		

In which 300 ms was the time spent in window segmentation, 7 ms for feature extraction, 9 ms recognition time. However, the Response time may be further reduced by analysing the pipeline efficiency when used with sEMG sensors having a higher sampling frequency. Table 2 summarized each component of the Response time for the proposed pipeline.

The plots in the Fig. 18 highlights the scalability of a classification model built using our proposed pipeline. The plot on the left describes the time taken(fit_times) in unit to train the LDA classifier on different number of samples while the plot on right describes the accuracy score archived with respect to the time(fit_times). Considering the two plot we can deduce that in less than 10000 samples the model achieves its maximum score which doesn't

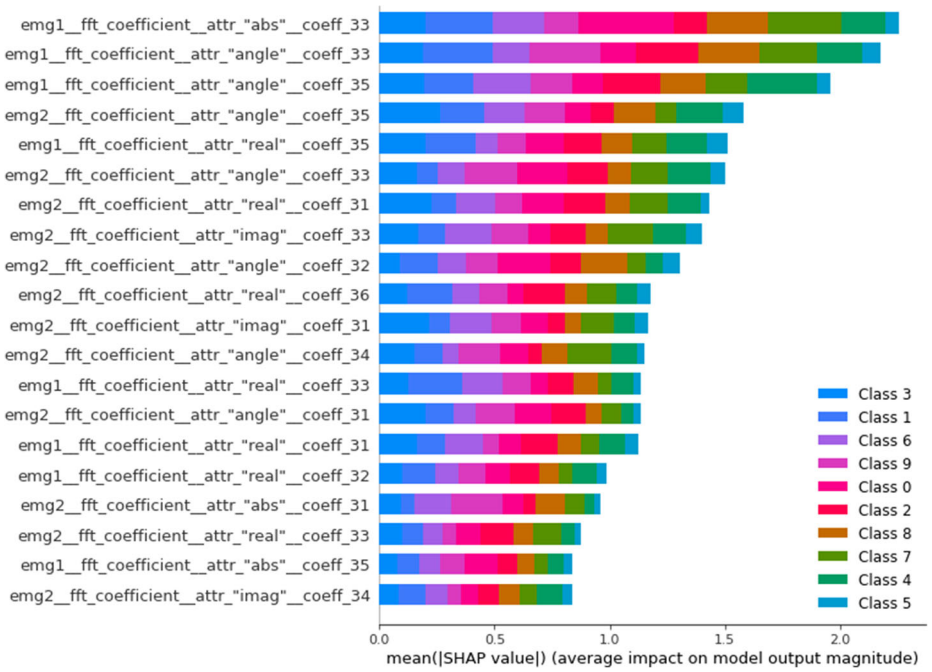

Fig. 16 Classwise feature importance illustrated using SHAP

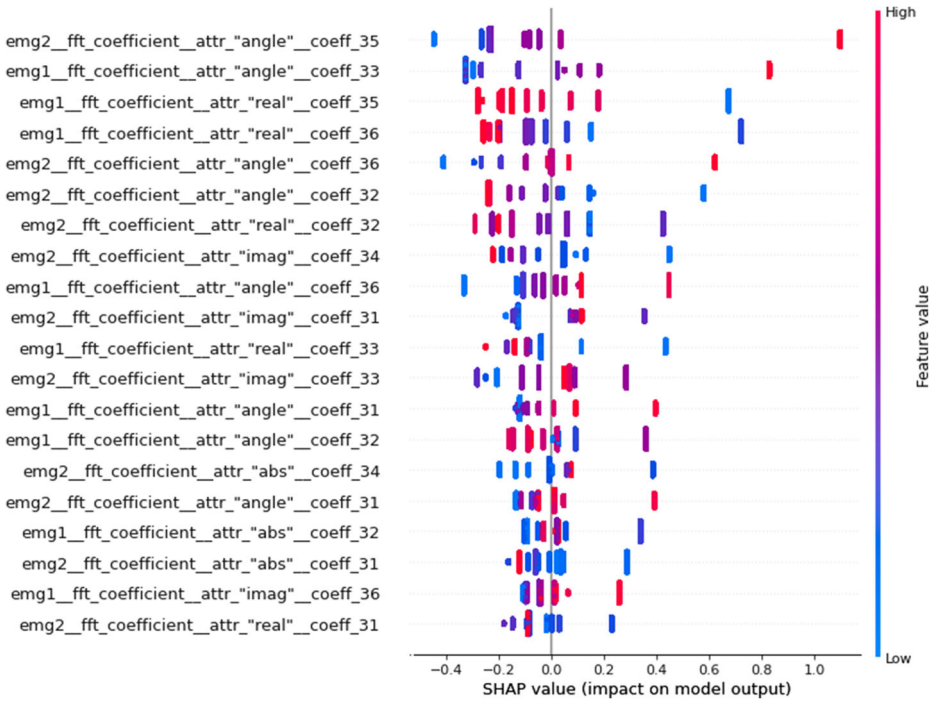

Fig. 17 Figure illustrating the impact of feature values on the model performance while predicting class 1 of the ASL dataset

varying largely with increasing the number of samples while training. This information can be used while adding the new gestures to the recognition model, which is not present during training. Also, the non varying accuracy scores suggests the robustness of the pipeline.

Moreover, a separate analysis for scalability of the pipeline was performed using the ASL-24 dataset. The ASL gestures for the dataset was deliberately increased up-to 24 different gestures. These 24 gestures corresponds to 24 ASL manual alphabets mainly using while finger-spelling ASL words. Using ten fold cross validation the model achieves the average classification accuracy of 99.91% and MCC score of 0.99. The Fig. 19 illustrate the MCC score achieved with respect to the iterations of the Catboost algorithm. Initially, the MCC scores lies in the range of (0.25-0.5). However, its reaches to maximum within 50 iterations. Morover, the Fig. 20 shows the model efficiency by plotting "Multi-Class" loss function against the iteration performed by Catboost algorithm. The validation and learning curve reaches to zero within 400 iterations. Both the learning and validation curve for all the ten folds are almost overlapping and smooth in nature suggesting the lower error produced by the model while predicting class label.

Table 2 Response time of the proposed pipeline to identify a single ASL gesture

Window segmentation (T_{w_s})	Feature Extraction (T_{fe})	Recognition Time (T_{Recog})	Response Time (T_{w_s}+T_{fe}+T_{Recog})
300 ms	7 ms	9 ms	316 ms

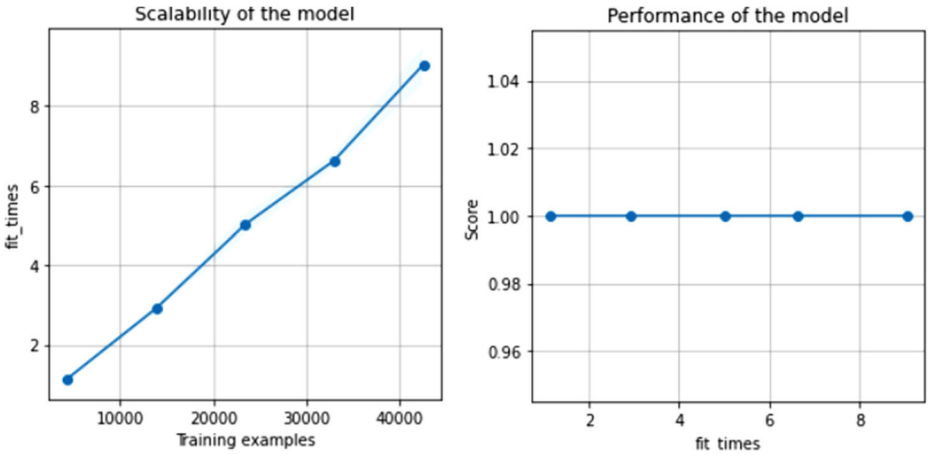

Fig. 18 The figure illustrating the time required to train various number of samples using LDA

4.0.1 Statistical analysis

We conducted a comprehensive statistical analysis of the obtained experimental results to validate their correctness, hence the performance of our proposed machine learning pipeline. Statistical hypothesis tests were applied to ensure that the results obtained were actual and not produced due to statistical fluke. Specifically, these analyses were performed with two primary objectives. First, to prove the statistical significance of the features selected using our proposed ensemble feature selection method. Second, to establish the correctness of the classification algorithm results by using the 5*2 cv t-test.

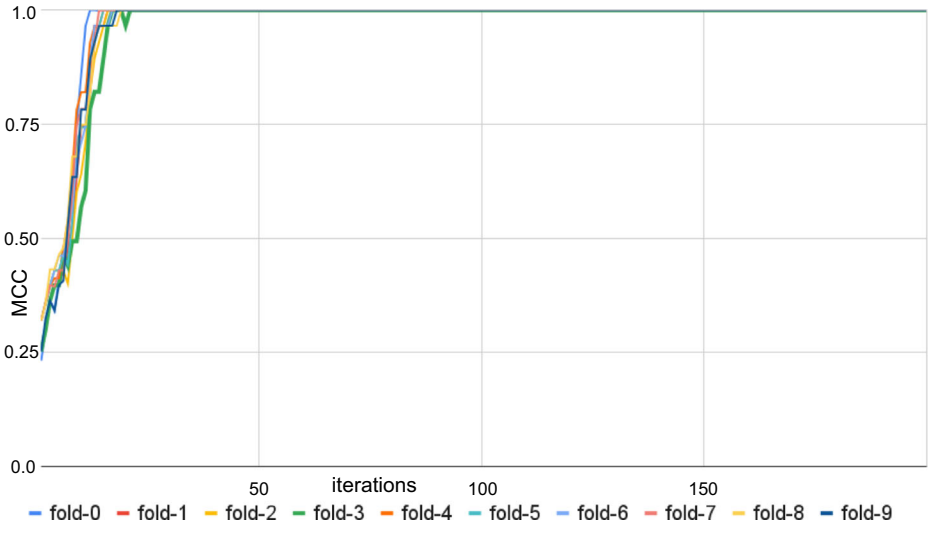

Fig. 19 MCC score plotted against number of Catboost iterations for each of 10 fold cross validation for ASL-24 dataset

Table 3 The selected features and the corresponding p-values

Channel 1 feature	p-value	Channel 2 feature	p-value
Fft coefficient 31 (REAL)	8.372341E-12	fft coefficient 32 (REAL)	0
Fft coefficient 36 (REAL)	1.5706702E-101	fft coefficient 35 (REAL)	7.696407E-68
Fft coefficient 31(IMG)	1	fft coefficient 36 (REAL)	1.570670E-101
Fft coefficient 32 (IMG)	0	fft coefficient 32 (IMG)	0
Fft coefficient 34 (IMG)	0.001298	fft coefficient 35 (IMG)	4.8492671E-13
Fft coefficient 35 (IMG)	4.849267E-13	fft coefficient 31(ABS)	0
Fft coefficient 34 (ABS)	0	fft coefficient 32 (ABS)	0
Fft coefficient 35 (ABS)	0	fft coefficient 33 (ABS)	0
Fft coefficient 36 (ABS)	0	fft coefficien 35 (ABS)	0
Fft coefficient 31 (ANG)	0	fft coefficient 31 (ANG)	0
Fft coefficient 32 (ANG)	0	fft coefficient 35 (ANG)	0.000083
Fft coefficient 34 (ANG)	0	fft coefficient 36 (ANG)	1.283962E-56
Fft coefficient 35 (ANG)	0.000083		

Kruskal Wallis test was applied to statistically substantiate the significance of the features selected by our proposed ensemble method [60]. The Kruskal Wallis test is a computationally efficient statistical method and has been frequently used in literature to determine features with significant discriminatory ability. Sharma and Pachauri [88] use the Kruskal Wallis statistical test to validate the discriminative ability of their proposed method to classify epileptic seizures and seizure-free signals. In similar work, Khan et al. [4] applied the Kruskal Wallis test to validate the features with higher discriminative information for the face recognition task. The features with p-values near zero were considered discriminative face features and claimed to produce a high recognition rate. In [66] the author claims to find significant features for the classification task when the p-value lies near 0 (p-value$<$0.5).

The Kruskal Wallis test is a non-parametric method. It mainly validates a null hypothesis that the medians of the groups/features used are similar and helps find whether samples originated from the same distribution. The H -statistics and the p-value is applied to determine the acceptance or rejection of the null hypothesis. We validated the distribution of the selected features using the Shapiro Wilk test before applying the Kruskal Wallis test. In our experiment, the Kruskal Wallis test rejected the null hypothesis for the ASL digit dataset with a p-value being less than 0.001.

However, the Kruskal Wallis hypothesis test doesn't reveal many details on the group/features that differ. A Post hoc test is required to compare the pairwise feature's significance. As the selected features were non-parametric, we chose the Dunn's test as a post hoc test [25]. In addition, Bonferroni adjustment was applied to deal with the cumulative Type I error or alpha inflation while using the Dunns test [23]. Bonferroni controls the Type I error by simply calculating a new pairwise alpha to keep the familywise alpha value at any specified value.

For the Dunns test, the resultant p values for all the pairs of features were reported to be less than 0.001. The Table 3 shows the pairwise p-values for the "emg1 fft coefficient 31 (IMG)" wrt to all the other features after Bonferroni adjustment. Figure 21 shows the schematic diagram of the statistical tests performed to validate the obtained results

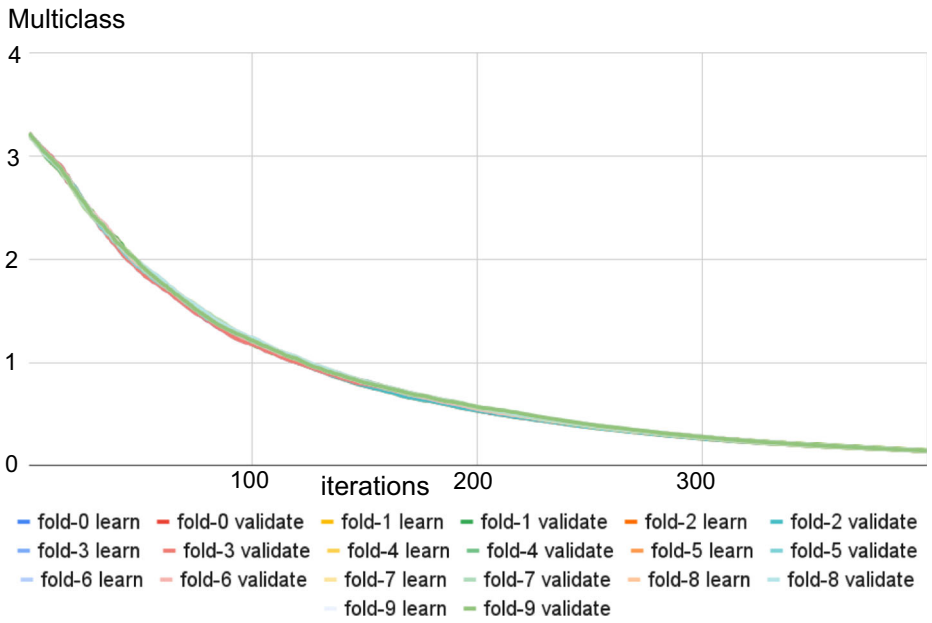

Fig. 20 Figure illustrating the Mutlticlass loss function using 10 fold cross validation(ASL-24 dataset)

4.0.2 5*2 CV t-test

To statistically validate the performance of our classification model, we used the 5*2 cv t-test proposed by Dietterich [22]. This statistical approach applies a t-test on five iterations of 2-fold cross-validation. 5*2 CV t-test can efficiently distinguish the classification algorithm's performance on a single dataset with an acceptable Type I error rate as compared to other statistical methods [22].

The use of the 5*2 cv t-test was based on our heuristics. If two different instances of the classifier, C_1, and C_2, provide statistically similar results on our selected feature set, then such results suggest the correctness and generalizability of our proposed pipeline. The two instances were made by assigning different hyperparameters to the classifier. C_1 corresponds to an instance of the classifier with default hyperparameters.In contrast, C_2 corresponds to the classifier instance with tuned hyper-parameters which provided the best average classification accuracy on the ASL 10 dataset.

The 5*2 cv t-test randomly partitioned the dataset into two equal-size sets, X_i(train set) and $\bar{X}_i$(test set), repeatedly for five iterations. In every iteration, each of the two classifiers,C_1 and C_2, are trained using the training set X_i and evaluated on the test set $\bar{X}_i$ and vice versa. This training and testing pattern generates the four error estimates $p_{C_1}^{(1)}$, $p_{C_2}^{(1)}$, $p_{C_1}^{(2)}$, $p_{C_2}^{(2)}$ in a single iteration. These error estimates are used to calculate the two performance measures($p_i^{(1)}$ and $p_i^{(2)}$) which are described as

$$p_i^{(1)} = p_{C_1}^{(1)} - p_{C_2}^{(1)} \tag{5}$$

$$p_i^{(2)} = p_{C_1}^{(2)} - p_{C_2}^{(2)} \tag{6}$$

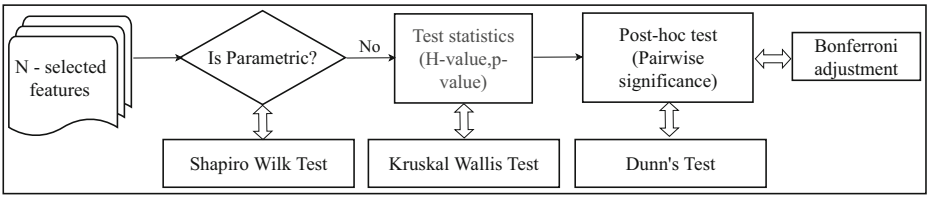

Fig. 21 The schematic diagram describing the statistical test performed to validate the results

With, (5) and (6) the variance can be calculated as $s_i^2 = (p_i^{(1)} - \bar{p}_i)^2 + (p_i^{(2)} - \bar{p}_i)^2$ where, $\bar{p}_i = (p_i^{(1)} + p_i^{(2)})/2$. Using these variables the 5*2cv t-statistics is described as follows:

$$\bar{t} = \frac{p_1^{(1)}}{\sqrt{\frac{1}{k}\sum_{i=1}^{k} s_i^2}} \tag{7}$$

Where, $k = 5$, $p_1^{(1)}$ is the error estimate of the first iteration and s_i^2 is the variance evaluated at i-th iteration. For our experiment, p-value was reported to be 0.96 (p-value>0.05), which suggests the performance of the two instance of the classifier are similar.

5 Discussion

This paper presented a reliable and accurate machine learning pipeline for sEMG based hand gesture recognition utilizing an optimal number of sensors 8-sEMG channels for (ASL-24) and 2-sEMG channels for (ASL-10) dataset. The result obtained highlights the application of Fourier transform as a sufficient feature extraction technique for classifying ASL and other similar hand gestures involving combined or single finger movements. The FFT coefficients evaluated for sEMG signals are complex numbers and can be decomposed into real (REAL), imaginary (IMAG), absolute (ABS), and angle (ANG) components. These individual components can be treated as a separate feature for recognizing the sEMG based gestures. The sEMG signals contain frequency in the range of 0-500 Hz. For our collected ASL gesture dataset, while performing classification, the FFT coefficients corresponding to the lower frequency component of sEMG signals show a more significant contribution than the higher frequency components. When treated as features, these lower frequency FFT components as a group exhibit sufficient information to identify effectively various hand gestures involving finger movements.

In the proposed pipeline, ensemble feature selection method helps list the best representative feature set among the extracted features for the classification task. The ensemble is performed in two steps to deal with the larger feature space. At the initial level, the feature space is reduced using the wrapper-based methods (ANOVA, CHI2, ReliefF, MUTUAL INFO). While the next level merges these selected features in a final selected set using the ranking, feature-feature and feature-class label correlation information. The correlation-based approach helps to obtain optimal number of highly significant non-correlated features in the final selected feature subset.

The selected features subset through the ensemble method provides sufficient discriminatory property/information to distinguish different ASL dataset classes accurately. A 2D projection of selected features for each class of the ASL gestures using a t-sne plot is shown in Fig. 13. The instances associated with varying ASL classes are used to form the data

points of the t-SNE plot. These datapoint associated with each ASL class is projected in lower dimension space, forming different clusters. The resultant least overlapping clusters for each class suggest the discriminatory ability of the resulting feature set, which eventually helps machine learning classifiers attain higher classification accuracy. While converting N-D feature space to 2-D feature space, the t-SNE algorithm ensures that the two points in a close distance in a higher dimension should remain close in generated lower dimension mapping. So two clusters representing different classes being mapped in a non-overlapping fashion in common feature space such as t-SNE suggest higher class separability. Similar observation about the selected features was made while applying the pipeline on the benchmark dataset (Ninapro database 5 exercise A). The resultant FFT coefficients corresponding to lower frequencies were able to classify the twelve different gestures involving finger movements with higher accuracy. When plotted in lower dimension with t-sne, these features too resulted in the least overlapping clusters supporting the inference made for our collected ASL gesture dataset.

Statistical tests performed validated the correctness of the experimental results. Moreover, the Kruskal Wallis test statistically substantiated the significance of the features subset selected by our proposed ensemble method. The reported p-value was less than 0.001, suggests the higher significance of the each of the selected features on the classification results. Moreover, the post hoc test cross validated the Kruskal Wallis results. In addition, the 5*2 cv t-test statistically validated the classification accuracy of the pipeline and ensures that the obtained result are not produced due to statistical fluke.

The reported efficiency of the FFT coefficient can be supported by the fact that similar observations about the FFT coefficients were analyzed in various other related works. In [26], the author found the Fast Fourier Transform coefficients as an efficient means for developing a sEMG signals-based camera control. Using Simple Principal Component Analysis, they achieved a classification accuracy of 81%. Similarly, in [72], the authors claim that the use of FFT is beneficial in identifying different leg movements, and in [37], FFT was used to classify neuromuscular disorders involving sEMG signals. However, the capability of these sEMG signals features need to further investigated in detail on dataset wuth larger number of ASL gestures.

Various Machine learning approaches have been applied to build an accurate and reliable classifier for ASL gesture recognition. As mentioned in the Related work, these methods deployed various sensors, feature extraction, and feature selection techniques to improve the classification accuracy. In Table 4, organizing the relevant information about the prominent research works for the sEMG based ASL recognition and our proposed work, we have tried to present the significance of our method for ASL gesture recognition. In order to compare the two methods, All the experiments should be performed and evaluated in similar circumstances. However, most of the dataset used in these research works are not publicly available, which prevents such a comparison.

However, a glance at the table's evaluation metrics highlights that using a machine learning classification pipeline based on Fourier Transform is an efficient means to obtain sufficient and acceptable results for recognizing ASL gestures involving finger movements.

Utilizing two sEMG channels, our proposed pipeline achieved a classification accuracy of 99.99% for 10 ASL gestures(ASL-10). Also, the integrated feature selection applied helps obtain the optimal number of features (25 features) for the classification tasks. The efficiency of the proposed model is generalized and has been validated on a dataset having a comparatively larger number of subjects and the number of samples compared to other research work. Considering the use of an optimal number of sensors, among the various

Table 4 Summary of the related work for sEMG sensor based ASL recognition

References	Different types of sensors used	No of samples	No of gestures	No of subjects	No of channels	Classification accuracy
Savur and Sahin [86]	1	1040	26	10	8	61.04%
Fatmi et al. [27]	4	26000	13	3	26	93.79%
Paudyal et al. [71]	3	–	20	10	34	97.72%
Wu et al. [101]	3	3000	40	4	10	95.94%
Wu et al. [99]	3	24000	80	4	10	85.24%-96.16%
Kosmidou et al. [53]	1	180	9	–	2	97.7%
Savur and Sahin et al. [85]	1	2080	26	–	8	91.1%
Taylor [92]	3	–	20	–	15	94%-98%
Proposed method (ASL-10)	1	53000	10	20	2	99.99%
Proposed method (ASL-24)	1	83000	24	10	8	99.91%

research work listed in the Table 4, only the authors [53] tried to use a reduced number of sensors (2 sEMG channels). Compared to this work, the proposed model achieves better classification accuracy and is more robust as they have only used 180 samples for model validation.

To analyze the computational overhead of a real-time recognition system built using our proposed pipeline, we tried to approximate end to end processing delay of such a system. The end-to-end delay can be approximated by calculating the individual delay of all the major components of an online classification system. Mainly an online classification system consists of various modules, including data acquisition (buffer storage)& segmentation (windowing), feature extraction, and class label prediction using pre-trained model [85]. We quantified the summation of each component's time delay in a single term called response time. Table 2 highlights the value of the response time for the system. All the values are evaluated on the workstation with Intel Core i7 having 2.9 GHz frequency with 16 GB RAM. The resultant processing delay is approximated as 316 ms, which is acceptable for real-time recognition.

Various approaches utilizing machine learning (ML) methods have been proposed to build an accurate and reliable classifier for sEMG based ASL gesture recognition. These ML-based approaches mainly deployed various sensors, extract relevant features for sEMG signals and then apply machine learning classification algorithms. Some of the most popular baseline approaches for sEMG-based recognition include Support Vector Machine (SVM), Hidden Markov model (HMM), Artificial neural network (ANN), linear discriminant classifier, etc. Savur et al. [86] uses the SVM classifier and achieves 61.04% classification accuracy on a multi-user dataset. While Fatmi et al. [27] compared the three machine learning baseline approaches (ANN, SVM, and HMM) for the ASL recognition task. The author claims that ANN achieves an overall higher accuracy of 93.79% than the other baseline approaches. (SVM 85.56% and HMM 85.90%).Meanwhile, Wu et al. [101] evaluated the Decision tree, LibSVM, Nearest Neighbour, and Naïve Bayes machine learning algorithm for ASL gesture recognition. Using an additional inertial sensor, the authors achieved a classification accuracy of 95.94%.

Dynamic Time Warping(DWT) [63] measures similarity between two temporal sequences and has also been commonly used for ASL recognition. Paudyal et al. [71] uses the DWT for classifying 20 ASL gesture. Using two additional sensors and sEMG,

they could attain 97.72% classification accuracy. However, for building a real-time ASL sign recognition system, comparatively higher recognition accuracy is required. Considering the performance of the baseline approaches in the research work discussed above, we tried to explore the newer algorithm for the classification task. The Catboost algorithm, with the ensemble feature selection approach, achieved a classification accuracy of 99.91% for a comparable number of gestures (24 ASL manual alphabets). The reported accuracy highlights the efficiency our method over the other baseline approaches.

Table 4, provides a comparison of prominent sEMG-based ASL recognition research works with our proposed work. The tables summarizes the performance of each research work highlighting the number of different type of sensors used, the subject involved in data collection, the dataset size, number of ASL gestures, number of sensors channel used and the performance metrics. For any two methods to be compared correctly, all related experiments must be performed and evaluated under similar conditions. But most of the datasets used in these research works are not publicly available, which prevents such a comparative analysis. However, a glance at the table's evaluation metrics suggests that our proposed pipeline perform better or comparable than state of art methods for ASL recognition task.

Considering the use of an optimal number of sensors, with two sEMG channels, our proposed pipeline achieved a classification accuracy of 99.99% for 10 ASL gestures . Also, the ensemble feature selection method helps obtain the optimal number of features (25 features) for the classification tasks. Among the various research work listed in the Table 4 only the authors in [53] tried to use a reduced number of sensors (2 sEMG channels). Compared to this work, our proposed model achieves better classification accuracy and is more robust as they have only used 180 samples for model validation.

Moreover, the proposed pipeline presented an efficient scalability while recognizing a increased number of ASL gestures. When applied on dataset consisting 24 manual alphabets gestures (ASL-24) it achieves the average classification accuracy of 99.91%. Comparing with the state of art methods (Table 4) the reported accuracy is highest for a approach using optimal number of sEMG sensors. The proposed pipeline with ASL-24 dataset, perform better when to compared with the work of Fatmi et al. [27], Paudyal et al. [71] , Taylor [92], Kosmidou et al. [53], Savur et al. [85]. We cannot directly compare classification accuracy with Wu et al. [99] and Wu et al.[101] as both of them have used a greater number ASL gestures. However, the performance of our approach is comparable to these two methods as they have used greater number of sensors and have validated their model on dataset on comparatively containing lesser number of samples. On the other hand our model is more generalizable as have been validated on dataset with larger number of people. For Savur et al. [86] and Savur et al. [85] attain a lower classification accuracy with comparable number of ASL gestures.Based on above comparison we can conclude that our proposed pipeline is better suited for developing cost effective SLRS using optimal number of sensors.

The use of sEMG sensors has an advantage over the other sensor-based ASL recognition system as these sensors are easily available in wearable wireless modules and are less susceptible to maintaining a line of sight with the receiver attached to the recognition module. This can increase the ease of use as compared to other sensor-based approaches. In addition, the use of sEMG sensors can provide a significant role even if the signer is suffering from upper limb amputation(Finger amputations). Our proposed pipeline is more interpretable and provides comparable classification accuracy than the deep learning models used for a similar gesture recognition task. Most deep learning pipelines are deployed end-to-end, implicitly extracting the features from the raw signals, giving less insight about the features

used. Thus a user-friendly, cost-effective, and reliable sEMG based SLRS system for ASL can be built using our proposed pipeline.

The Myo armband used in the experiments can communicate with Android smartphones through various interfaces. Effort would be made to implement a cost-effective ASL recognition system using the smartphone, Myo armband, and the proposed model in our future work. The reported experimental results further can be utilized field of robotics/prosthetics involving replicating, decoding, or identifying the finger or hand gesture movements to improve prosthetic control. In addition, can be used to improve precision and control while building sEMG based Human-computer Interaction applications [3].

5.1 Limitations and challenges

Despite providing better classification accuracy for the ASL gesture recognition task, our proposed machine learning pipeline poses certain limitations, which are as follows:

a) The accuracy of the sEMG-based gesture recognition is affected by various factors such as the sensor placement, intra & inter-day data collection, muscle fatigue, and skin impedance [20, 48, 110]. However, during our experiments, we followed a strict protocol to minimize the effect of these variables. The sensor's placements and distance between these sensors were strictly monitored, data collection was performed in an intraday scenario, and various short sessions were separated by relaxation time to deal with muscle fatigue. In addition, the conductive gel was administered over the sensor placement area to deal with skin impedance. Evaluating the proposed machine learning pipeline in an environment without such protocols may adversely affect the overall recognition accuracy.
b) The ASL gestures considered for recognition tasks are primarily static in nature. Whereas the ASL dictionary contains different dynamic words (consists of various hand motions). The proposed machine learning pipeline may behave differently when applied to these dynamic gestures.
c) Non-manual markers (facial expressions) are an essential part of ASL communication. A recognition module based on sEMG signals used cannot capture this information or differentiate between two ASL signs involving Non-manual markers. Hence, the proposed pipeline is ineffective for ASL gestures heavily dependent on Non-manual markers.
d) The reported classification accuracy is based on evaluating isolated ASL gestures. However, during real-time ASL communications, a signer generates continuous gestures. To effectively work in a continuous gesture scenario, a recognition module must be able to identify the presence and duration of a gesture performed. So for real-time use, the proposed machine learning pipeline has to be integrated with algorithms that can predict the presence and duration of the ASL gesture.

Despite the advancements in technology such as sensor fabrications, activity recognition, and deep learning algorithms, a commercial, cost-effective ASL system is not available. Around 70 million deaf people uses nearly 300 sign languages worldwide [11]. The availability of a cost-effective accurate SLRS could help improve the day-to-day standard of living of these people. One of the major challenges in SLRS is maintaining acceptable recognition accuracy in real-time without increasing the overall system complexity and cost [5]. Recognition accuracy can be improved by integrating multiple or complex hybrid sensors, which eventually affects the price. Hence a tradeoff between the number of sensors and

cost must be maintained. Developing accurate machine learning models, channel optimization algorithms, cost-effective hybrid sensors, and relevant compatible software are some of the other major challenges which need to be dealt with in the SLRS field.

6 Conclusion and future work

This paper presented an sEMG-based machine learning pipeline for Sign language Recognition and accessed its validity on two newly collected datasets consisting of sEMG signals corresponding to different ASL gestures. The proposed pipeline is also validated on a benchmark dataset(Ninapro Database 5), frequently used to access sEMG-based gesture recognition methods. Validation of the pipeline is performed in terms of average classification accuracy, MCC, AUC, response time, and scalability. Experimental results show that our pipeline achieved an excellent classification accuracy of 99.91% and other performance parameters compared to recently published research work showing the proposed work's novelty. The statistical test such as Kruskal Wallis and 5*2 cv t-test confirm the statistical significance of the reported results.

Our experiments concluded that properly curating the quality of the extracted features for sEMG signals can result in improved performance of sensor-based Sign Language Recognition system and other similar hand gesture analysis tasks. Also, the FTT analysis plays a significant role in sEMG based gesture recognition model. Using the various FFT coefficients as features can help improve the performance of a gesture recognition system involving finger movements. Sensor-based systems, less susceptible to lighting conditions, can be treated as a better alternative to visual methods for building accurate sign language recognition systems.

In future, our proposed recognition pipeline can be used to develop an advanced sign language recognition system capable of recognizing complete ASL sentences. For sentence-level recognition, individual sign words initially needs to be identified by separately using a machine learning classification module similar to our proposed pipeline. Generating ASL sentences requires continuous hand motion. Hence for correct identification of ASL words, an automatic gesture presence detection and segmentation algorithm need to be developed and integrated. Further, being correctly identified these individually sign words can be put in sequential order to reconstruct the sentence meaning. The accuracy of the model can be further improved by integrating it with Natural Language Processing (NLP) models such as N-gram(customized for ASL grammar). Such a sentence-level ASL recognition system would significantly impact ASL signers while learning or expressing sign languages. Despite the advancement in HCI, an accurate, cost-effective SLRS is not available commercially. The paper demonstrates the feasibility of using the sEMG-based recognition model in building such SLRS.

Acknowledgements We thank Dr. Shiru Sharma, Associate Professor, School of Biomedical Engineering IIT BHU, for providing the MyoArmband for data collection. We also thank Dr. Alok Prakash, School of Biomedical Engineering IIT BHU, for help in deciding the experimental protocol for data collection.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interests None of the authors have any conflict of interest to declare.

References

1. Aboy M, Hornero R, Abásolo D, Álvarez D (2006) Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis. IEEE Trans Biomed Eng 53(11):2282–2288
2. Ahmed MA, Zaidan BB, Zaidan AA, Salih MM, Lakulu MMB (2018) A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017. Sensors 18(7):2208
3. Ahsan MR, Ibrahimy MI, Khalifa OO et al (2009) Emg signal classification for human computer interaction: a review. Eur J Sci Res 33(3):480–501
4. Ali Khan S, Hussain A, Basit A, Akram S (2014) Kruskal-wallis-based computationally efficient feature selection for face recognition. Sci World J, vol 2014
5. Anderson R, Wiryana F, Ariesta MC, Kusuma GP et al (2017) Sign language recognition application systems for deaf-mute people: a review based on input-process-output. Procedia Comput Sci 116:441–448
6. Atzori M, Gijsberts A, Castellini C, Caputo B, Hager A-GM, Elsig S, Giatsidis G, Bassetto F, Müller H (2014) Electromyography data for non-invasive naturally-controlled robotic hand prostheses. Sci Data 1(1):1–13
7. Barbhuiya AA, Karsh RK, Jain R (2021) Cnn based feature extraction and classification for sign language. Multimed Tools Appl 80(2):3051–3069
8. Batista GE, Keogh EJ, Tataw OM, De Souza VM (2014) Cid: an efficient complexity-invariant distance for time series. Data Min Knowl Disc 28(3):634–669
9. Battison R (1978) Lexical borrowing in american sign language
10. Bheda V, Radpour D (2017) Using deep convolutional networks for gesture recognition in american sign language. arXiv:1710.06836
11. Bin Munir M, Alam FR, Ishrak S, Hussain S, Shalahuddin M, Islam MN (2021) A machine learning based sign language interpretation system for communication with deaf-mute people. In: Proceedings of the XXI international conference on human computer interaction, pp 1–9
12. Blum AL, Langley P (1997) Selection of relevant features and examples in machine learning. Artif Intell 97(1-2):245–271
13. Cardenas EJE, Chavez GC (2020) Multimodal hand gesture recognition combining temporal and pose information based on cnn descriptors and histogram of cumulative magnitudes. J Vis Commun Image Represent 71:102772
14. Chang Y-W, Lin C-J (2008) Feature ranking using linear svm. In: Causation and prediction challenge. PMLR, pp 53–64
15. Chen C-W, Tsai Y-H, Chang F-R, Lin W-C (2020) Ensemble feature selection in medical datasets: combining filter, wrapper, and embedded feature selection results. Expert Syst 37(5):12553
16. Chicco D, Jurman G (2020) The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC Genomics 21(1):1–13
17. Chowdhury RH, Reaz MB, Ali MABM, Bakar AA, Chellappan K, Chang T (2013) Surface electromyography signal processing and classification techniques. Sensors (Basel Switzerland) 13(9):12431–12466
18. Chuan C-H, Regina E, Guardino C (2014) American sign language recognition using leap motion sensor. In: 2014 13th International conference on machine learning and applications. IEEE, pp 541–544
19. Cooper H, Holt B, Bowden R (2011) Sign language recognition. In: Visual analysis of humans. Springer, pp 539–562
20. Day S (2002) Important factors in surface emg measurement. Bortec Biomed Ltd Pub:1–17
21. De la Rosa R, Alonso A, Carrera A, Durán R, Fernández P (2010) Man-machine interface system for neuromuscular training and evaluation based on emg and mmg signals. Sensors (Basel Switzerland) 10(12):11100–11125
22. Dietterich TG (1998) Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computat 10(7):1895–1923
23. Dinno A (2015) Nonparametric pairwise multiple comparisons in independent groups using dunn's test. Stata J 15(1):292–300
24. Dorogush AV, Ershov V, Gulin A (2018) Catboost: gradient boosting with categorical features support. arXiv:1810.11363
25. Dunn OJ (1964) Multiple comparisons using rank sums. Technometrics 6(3):241–252
26. Erkilinc MS, Sahin F (2011) Camera control with emg signals using principal component analysis and support vector machines. In: 2011 IEEE international systems conference. IEEE, pp 417–421

27. Fatmi R, Rashad S, Integlia R (2019) Comparing ann, svm, and hmm based machine learning methods for american sign language recognition using wearable motion sensors. In: 2019 IEEE 9th annual computing and communication workshop and conference (CCWC)
28. Fels SS, Hinton GE (1993) Glove-talk: a neural network interface between a data-glove and a speech synthesizer. IEEE Trans Neural Netw 4(1):2–8
29. Feng Y, Uchidiuno UA, Zahiri HR, George I, Park AE, Mentis H (2021) Comparison of kinect and leap motion for intraoperative image interaction. Surg Innov 28(1):33–40
30. Ferri C, Hernández-Orallo J, Modroiu R (2009) An experimental comparison of performance measures for classification. Pattern Recognit Lett 30(1):27–38
31. Friedrich R, Siegert S, Peinke J, Siefert M, Lindemann M, Raethjen J, Deuschl G, Pfister G et al (2000) Extracting model equations from experimental data. Phys Lett A 271(3):217–222
32. Garcia B, Viesca SA (2016) Real-time american sign language recognition with convolutional neural networks. Convolutional Neural Netw Vis Recognit 2:225–232
33. Genuer R, Poggi J-M, Tuleau-Malot C (2015) Vsurf: an r package for variable selection using random forests. R Journal 7(2):19–33
34. Gomez-Donoso F, Orts-Escolano S, Cazorla M (2019) Accurate and efficient 3d hand pose regression for robot hand teleoperation using a monocular rgb camera. Expert Syst Appl 136:327–337
35. Goswami T, Javaji SR (2021) Cnn model for american sign language recognition. In: ICCCE 2020. Springer, pp 55–61
36. Grandini M, Bagli E, Visani G (2020) Metrics for multi-class classification: an overview. arXiv:2008.05756
37. Güler NF, Koçer S (2005) Classification of emg signals using pca and fft. J Med Syst 29(3):241–250
38. Guo D, Zhou W, Li H, Wang M (2018) Hierarchical lstm for sign language translation. In: Proceedings of the AAAI conference on artificial intelligence, vol 32
39. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3(Mar):1157–1182
40. Haq AU, Zhang D, Peng H, Rahman SU (2019) Combining multiple feature-ranking techniques and clustering of variables for feature selection. IEEE Access 7:151482–151492. https://doi.org/10.1109/ACCESS.2019.2947701
41. Hoque N, Singh M, Bhattacharyya DK (2018) Efs-mi: an ensemble feature selection method for classification. Complex Intell Syst 4(2):105–118
42. Hudgins B, Parker P, Scott RN (1993) A new strategy for multifunction myoelectric control. IEEE Trans Biomed Eng 40(1):82–94
43. Isaacs J, Foo S (2004) Hand pose estimation for american sign language recognition. In: Thirty-sixth southeastern symposium on system theory, 2004. Proceedings of the. IEEE, pp 132–136
44. Jain A, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. IEEE Trans Pattern Anal Mach Intell 19(2):153–158. https://doi.org/10.1109/34.574797
45. Jones E, Oliphant T, Peterson P et al (2001) SciPy: open source scientific tools for python. http://www.scipy.org/. Accessed 10 June 2022
46. Jurman G, Riccadonna S, Furlanello C (2012) A comparison of mcc and cen error measures in multi-class prediction
47. Kadhim RA, Khamees M (2020) A real-time american sign language recognition system using convolutional neural network for real datasets. TEM J 9(3):937
48. Kanoga S, Kanemura A, Asoh H (2020) Are armband semg devices dense enough for long-term use?—sensor placement shifts cause significant reduction in recognition accuracy. Biomed Signal Process Contr 60:101981
49. Kerber F, Schardt P, Löchtefeld M (2015) Wristrotate: a personalized motion gesture delimiter for wrist-worn devices. In: Proceedings of the 14th international conference on mobile and ubiquitous multimedia, pp 218–222
50. Khan SM, Khan AA, Farooq O (2019) Selection of features and classifiers for emg-eeg-based upper limb assistive devices—a review. IEEE Rev Biomed Eng 13:248–260
51. Kleiman R, Page D (2019) Aucμ: a performance metric for multi-class machine learning models. In: International conference on machine learning. PMLR, pp 3439–3447
52. Koller O (2020) Quantitative survey of the state of the art in sign language recognition. arXiv:2008.09918
53. Kosmidou VE, Hadjileontiadis LJ, Panas SM (2006) Evaluation of surface emg features for the recognition of american sign language gestures. In: 2006 International conference of the IEEE engineering in medicine and biology society, pp 6197–6200. https://doi.org/10.1109/IEMBS.2006.259428
54. Kuroda T, Tabata Y, Goto A, Ikuta H, Murakami M et al (2004) Consumer price data-glove for sign language recognition. In: Proceeding ICDVRAT, pp 253–258

55. Lee CK, Ng KK, Chen C-H, Lau HC, Chung S, Tsoi T (2021) American sign language recognition and training method with recurrent neural network. Expert Syst Appl 167:114403
56. Li L, Jiang S, Shull PB, Gu G (2018) Skingest: artificial skin for gesture recognition via filmy stretchable strain sensors. Adv Robot 32(21):1112–1121
57. Liddell SK, Johnson RE (1989) American sign language: the phonological base. Sign Language Studies 64(1):195–277
58. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems 30. Curran Associates, Inc., pp 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf. Accessed 10 June 2022
59. Masood S, Srivastava A, Thuwal HC, Ahmad M (2018) Real-time sign language gesture (word) recognition from video sequences using cnn and rnn. In: Intelligent engineering informatics. Springer, pp 623–632
60. McKight PE, Najab J (2010) Kruskal-wallis test. corsini Encyclo Psychol:1–1
61. Mehdi SA, Khan YN (2002) Sign language recognition using sensor gloves. In: Proceedings of the 9th international conference on neural information processing, 2002. ICONIP'02., vol 5, pp 2204–22065. https://doi.org/10.1109/ICONIP.2002.1201884
62. Miften FS, Diykh M, Abdulla S, Siuly S, Green JH, Deo RC (2021) A new framework for classification of multi-category hand grasps using emg signals. Artif Intell Med 112:102005
63. Müller M (2007) Information retrieval for music and motion. Springer, vol 2
64. Munib Q, Habeeb M, Takruri B, Al-Malik HA (2007) American sign language (asl) recognition based on hough transform and neural networks. Expert Syst Appl 32(1):24–37
65. NIDCD (2021) American sign language. https://www.nidcd.nih.gov/health/american-sign-language. Accessed 26 May 2021
66. Nishad A, Upadhyay A, Pachori RB, Acharya UR (2019) Automated classification of hand movements using tunable-q wavelet transform based filter-bank with surface electromyogram signals. Futur Gener Comput Syst 93:96–110
67. Olsson JOS, Oard DW (2006) Combining feature selectors for text classification. In: Proceedings of the 15th ACM international conference on information and knowledge management, pp 798–799
68. Oz C, Leu MC (2005) Recognition of finger spelling of american sign language with artificial neural network using position/orientation sensors and data glove. In: International symposium on neural networks. Springer, pp 157–164
69. Oz C, Leu MC (2007) Linguistic properties based on american sign language isolated word recognition with artificial neural networks using a sensory glove and motion tracker. Neurocomputing 70(16-18):2891–2901
70. Oz C, Leu MC (2011) American sign language word recognition with a sensory glove using artificial neural networks. Eng Appl Artif Intell 24(7):1204–1213
71. Paudyal P, Banerjee A, Gupta SK (2016) Sceptre: a pervasive, non-invasive, and programmable gesture recognition technology. In: Proceedings of the 21st international conference on intelligent user interfaces, pp 282–293
72. Pires R, Falcari T, Campo AB, Pulcineli BC, Hamill J, Ervilha UF (2019) Using a support vector machine algorithm to classify lower-extremity emg signals during running shod/unshod with different foot strike patterns. J Appl Biomechan 35(1):87–90
73. Pizzolato S, Tagliapietra L, Cognolato M, Reggiani M, Müller H, Atzori M (2017) Comparison of six electromyography acquisition setups on hand movement classification tasks. PloS One 12(10):0186132
74. Poizner H, Tallal P (1987) Temporal processing in deaf signers. Brain Lang 30(1):52–62
75. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) Catboost: unbiased boosting with categorical features. In: Advances in neural information processing systems, pp 6638–6648
76. Pugeault N, Bowden R (2011) Spelling it out: real-time asl fingerspelling recognition. In: 2011 IEEE international conference on computer vision workshops (ICCV workshops). IEEE, pp 1114–1119
77. Rao GA, Syamala K, Kishore P, Sastry A (2018) Deep convolutional neural networks for sign language recognition. In: 2018 Conference on signal processing and communication engineering systems (SPACES). IEEE, pp 194–197
78. Rashid O, Al-Hamadi A, Michaelis B (2010) Utilizing invariant descriptors for finger spelling american sign language using svm. In: International symposium on visual computing. Springer, pp 253–263
79. Rastgoo R, Kiani K, Escalera S (2018) Multi-modal deep hand sign language recognition in still images using restricted boltzmann machine. Entropy 20(11):809
80. Rastgoo R, Kiani K, Escalera S (2020) Hand sign language recognition using multi-view hand skeleton. Expert Syst Appl 150:113336

81. Remeseiro B, Bolon-Canedo V (2019) A review of feature selection methods in medical applications. Comput Bio Med 112:103375
82. Rivera-Acosta M, Ruiz-Varela JM, Ortega-Cisneros S, Rivera J, Parra-Michel R, Mejia-Alvarez P (2021) Spelling correction real-time american sign language alphabet translation system based on yolo network and lstm. Electronics 10(9):1035
83. Rodríguez-Tapia B, Soto I, Martínez DM, Arballo NC (2020) Myoelectric interfaces and related applications: current state of emg signal processing–a systematic review. IEEE Access 8:7792–7805
84. Salo F, Injadat M, Moubayed A, Nassif AB, Essex A (2019) Clustering enabled classification using ensemble feature selection for intrusion detection. In: 2019 International conference on computing, networking and communications (ICNC). IEEE, pp 276–281
85. Savur C, Sahin F (2015) Real-time american sign language recognition system using surface emg signal. In: 2015 IEEE 14th international conference on machine learning and applications (ICMLA). IEEE, pp 497–502
86. Savur C, Sahin F (2016) American sign language recognition system by using surface emg signal. In: 2016 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 002872–002877
87. Schreiber T, Schmitz A (1997) Discrimination power of measures for nonlinearity in a time series. Phys Rev E 55(5):5443
88. Sharma R, Pachori RB (2015) Classification of epileptic seizures in eeg signals based on phase space representation of intrinsic mode functions. Expert Syst Appl 42(3):1106–1117
89. Simons EDMGF, Fennig CD (2021) Ethnologue: languages of the world. http://www.ethnologue.com. Accessed 26 May 2021
90. Starner T, Pentland A (1997) Real-time american sign language recognition from video using hidden markov models. In: Motion-based recognition. Springer, pp 227–243
91. Sun C, Zhang T, Bao B-K, Xu C (2013) Latent support vector machine for sign language recognition with kinect. In: 2013 IEEE international conference on image processing. IEEE, pp 4190–4194
92. Taylor J (2016) Real-time translation of american sign language using wearable technology
93. Too J, Abdullah A, Saad NM, Ali NM, Musa H (2018) A detail study of wavelet families for emg pattern recognition. Int J Electr Comput Eng (IJECE) 8(6):4221–4229
94. Van der Maaten L, Hinton G (2008) Visualizing data using t-sne. J Mach Learn Res, vol 9(11)
95. Wadhawan A, Kumar P (2019) Sign language recognition systems: a decade systematic literature review. Arch Computat Methods Eng:1–29
96. Wadhawan A, Kumar P (2020) Deep learning-based sign language recognition system for static signs. Neural Comput Appl 32(12):7957–7968
97. Wang H, Khoshgoftaar TM, Napolitano A (2012) Software measurement data reduction using ensemble techniques. Neurocomputing 92:124–132
98. Wattenberg M, Viégas F, Johnson I (2016) How to use t-sne effectively. Distill. https://doi.org/10.23915/distill.00002
99. Wu J, Sun L, Jafari R (2016) A wearable system for recognizing american sign language in real-time using imu and surface emg sensors. IEEE J Biomed Health Inform 20(5):1281–1290. https://doi.org/10.1109/JBHI.2016.2598302
100. Wu J, Tian Z, Sun L, Estevez L, Jafari R (2015) Real-time american sign language recognition using wrist-worn motion and surface emg sensors. In: 2015 IEEE 12th international conference on wearable and implantable body sensor networks (BSN). IEEE, pp 1–6
101. Wu J, Tian Z, Sun L, Estevez L, Jafari R (2015) Real-time american sign language recognition using wrist-worn motion and surface emg sensors. In: 2015 IEEE 12th international conference on wearable and implantable body sensor networks (BSN). https://doi.org/10.1109/BSN.2015.7299393, pp 1–6
102. Yeh C-CM, Zhu Y, Ulanova L, Begum N, Ding Y, Dau HA, Silva DF, Mueen A, Keogh E (2016) Matrix profile i: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In: 2016 IEEE 16th international conference on data mining (ICDM). Ieee, pp 1317–1322
103. Yu E, Cho S (2006) Ensemble based on ga wrapper feature selection. Comput Industr Eng 51(1):111–116
104. Zafrulla Z, Brashear H, Starner T, Hamilton H, Presti P (2011) American sign language recognition with the kinect. In: Proceedings of the 13th international conference on multimodal interfaces, pp 279–286
105. Zamani M, Kanan HR (2014) Saliency based alphabet and numbers of american sign language recognition using linear feature extraction. In: 2014 4th International conference on computer and knowledge engineering (ICCKE). IEEE, pp 398–403
106. Zhang J, Bi H, Chen Y, Wang M, Han L, Cai L (2019) Smarthandwriting: handwritten chinese character recognition with smartwatch. IEEE Internet Things J 7(2):960–970

107. Zhang Y, Gong D-W, Cheng J (2017) Multi-objective particle swarm optimization approach for cost-based feature selection in classification. IEEE/ACM Trans Computat Bio Bioinform 14(1):64–75. https://doi.org/10.1109/TCBB.2015.2476796
108. Zhao W (2016) A concise tutorial on human motion tracking and recognition with microsoft kinect. Sci China Inf Sci 59(9):1–5
109. Zheng M, Crouch M, Eggleston MS (2021) Surface electromyography as a natural human-machine interface: a review. arXiv:2101.04658
110. Zia ur Rehman M, Gilani SO, Waris A, Niazi IK, Slabaugh G, Farina D, Kamavuako EN (2018) Stacked sparse autoencoders for emg-based classification of hand motions: a comparative multi day analyses between surface and intramuscular emg. Appl Sci 8(7):1126

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Affiliations

Shashank Kumar Singh[1] · Amrita Chaturvedi[1]

Amrita Chaturvedi
amrita.cse@iitbhu.ac.in

[1] CSE Deptt., Indian Institute of Technology (BHU), Varanasi, 221005, U.P., India