Filipino sign language alphabet recognition using Persistent Homology Classification Algorithm (#99720)

First submission

Guidance from your Editor

Please submit by **3 Jul 2024** for the benefit of the authors (and your token reward) .

Structure and Criteria
Please read the 'Structure and Criteria' page for guidance.

Author notes
Have you read the author notes on the guidance page?

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files
Download and review all files from the materials page.

6 Figure file(s)
2 Latex file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- **Impact and novelty is not assessed.** Meaningful replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Filipino sign language alphabet recognition using Persistent Homology Classification Algorithm

Cristian Jetomo Corresp., [1] , **Mark Lexter De Lara** [1]

[1] Institute of Mathematical Sciences and Physics, College of Arts and Sciences, University of the Philippines Los Baños, College, Los Baños, Laguna, Philippines

Corresponding Author: Cristian Jetomo
Email address: cbjetomo@up.edu.ph

Deaf or hearing-impaired individuals have been facing problems in communicating with the normal hearing population. To cope with this communication gap, numerous sign languages have been developed, one of which is the Filipino Sign Language (FSL). Despite FSL being declared as the national sign language of the Philippines, most Filipinos still do not understand the language. Hence, machine learning techniques are leveraged to automate the interpretation process of signed gestures and the field of sign language recognition is developed. This paper extends this field by utilizing computational topology-based methods in performing Filipino sign language recognition (FSLR). Specifically, it aims to utilize Persistent Homology Classification Algorithm (PHCA) in classifying or interpreting static alphabet signed using FSL. The performance of PHCA is evaluated in comparison with widely used classifiers. Validation runs shows that PHCA performed at par with the other classifiers considered.

Filipino sign language alphabet recognition using Persistent Homology Classification Algorithm

Cristian B. Jetomo[1] **and Mark Lexter D. De Lara**[2]

[1]**Institute of Mathematical Sciences and Physics, College of Arts and Sciences, University of the Philippines Los Baños, College, Los Baños, Laguna, Philippines**
[2]**Institute of Mathematical Sciences and Physics, College of Arts and Sciences, University of the Philippines Los Baños, College, Los Baños, Laguna, Philippines**

Corresponding author:
Cristian Jetomo[1]

Email address: cbjetomo@up.edu.ph

ABSTRACT

Deaf or hearing-impaired individuals have been facing problems in communicating with the normal hearing population. To cope with this communication gap, numerous sign languages have been developed, one of which is the Filipino Sign Language (FSL). Despite FSL being declared as the national sign language of the Philippines, most Filipinos still do not understand the language. Hence, machine learning techniques are leveraged to automate the interpretation process of signed gestures and the field of sign language recognition is developed. This paper extends this field by utilizing computational topology-based methods in performing Filipino sign language recognition (FSLR). Specifically, it aims to utilize Persistent Homology Classification Algorithm (PHCA) in classifying or interpreting static alphabet signed using FSL. The performance of PHCA is evaluated in comparison with widely used classifiers. Validation runs shows that PHCA performed at par with the other classifiers considered.

INTRODUCTION

Deafness or hearing loss has been one of the top causes of disability. In 2018, the World Health Organization (WHO) estimates that there are currently around 466 million people with disabling hearing loss globally (Davis and Hoffman, 2019). WHO projects that by 2050, nearly 2.5 billion people will have some degree of hearing loss and at least 700 million people will require hearing rehabilitation. One major challenge faced by deaf individuals is communication, an essential component in human existence. This challenge prevents further development in a lot of crucial aspects including education, specifically in terms of performance of hearing-impaired individuals (Most, 2004) and in turn, employment (Cruz and Calimpusan, 2018).

To cope with the communication gap and limited social interaction, the emergence of sign languages across different countries has taken place. These languages are done through hand gestures, dynamic movements, facial expressions, body motion, and palm orientation to portray the corresponding meaning of the signed gesture more effectively (Rivera and Ong, 2018). Sign languages have been very effective in addressing communication barrier and thus, has been one of the primary means of communication for the deaf community. In the Philippine context, FSL is declared as its national sign language which highlights the recognition, promotion, and support to FSL in all transactions. Despite this, there is still a lack of formal implementation of such programs, furthering the gap between the normal hearing and hearing-impaired population in the country.

Sign language interpreters are individuals who translate signed gestures into words or phrases and vice versa. Their help is of great importance to bridging this gap issue. However, with the limited number of expert sign interpreters, there is still a need for better solutions. Hence, many researchers leveraged the use of machine learning (ML) techniques to automate the interpretation or recognition process of signed gestures, leading to the development of the research field called Sign Language Recognition (SLR).

Often, SLR studies focus on vision-based and sensor or glove-based approach. However, there is a limited number of SLR studies focusing on FSL (Cabalfin et al., 2012; Montefalcon et al., 2023, 2021; Oliva et al., 2018; Rivera and Ong, 2018; Sandjaja and Marcos, 2009), even less on studies that recognize dynamic FSL or dataset of FSL that captures motion [citations]. These limitations what this paper tries to address.

Topological data analysis (TDA) is a newly emerging field that harnesses techniques from computational topology to analyze data. It makes use of these concepts to extract shape or topological features, usually via Persistent Homology (PH), which is important for ML problems. A survey is conducted by Hensel et al. (2021) to review and synthesize the current state of the fuse of TDA and ML. The authors divided some of these applications into two parts: extrinsic and intrinsic approach. Extrinsic topological approach utilizes PH to obtain a representation of data in the form of persistence diagrams. These diagrams are converted into features, using either vector-based or kernel-based representations, which are then fed into the common ML models. On the other hand, intrinsic topological approach incorporates TDA in the ML model itself. This approach either includes topological information into the design of the model or applies topology to study and improve the model. One example of the intrinsic approach is the development of the novel supervised classifier called Persistent Homology Classification Algorithm (PHCA).

PHCA has been implemented on different variants of datasets and have shown to perform at par if not better than the majority of classical classifiers including Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), K Nearest Neighbors (KNN), Classification and Regression Trees (CART), and Random Forest (RF). This study aims to extend the capabilities of the topology-based classifier by using FSL alphabet as datasets. Specifically, this paper aims to utilize PHCA to classify 10,800 images of FSL alphabet. Features are extracted using MediaPipe Hands (Zhang et al., 2020), a pipeline that detects and tracks the coordinates of certain landmarks of the hands.

The remainder of this paper is organized as follows. Section 2 presents related studies focusing on FSLR, revealing current state of the FSL. Section 3 elaborates on the important methods and techniques used in the study. Section 4 provides the presentation of results and discussion on these findings. Finally Sections 5 and 6 highlights the conclusion, some recommendations, and possible further researches of the authors on FSLR.

RELATED WORKS

The Philippine Deaf Research Center and the Philippine Federation of the Deaf states that the main components to extract from sign languages are hand shape, hand location, palm orientation and movements, and non-manual signals such as facial expressions (Center and of the Deaf, 2004). These components are the highlight of the feature extraction techniques being used in SLR. The two main established approach for SLR are sensor-based and visual-based. A hybrid approach is also being developed recently but will not be discussed in this paper due to limited FSLR works on it. Research works that do not use FSL as datasets are explored but are not discussed in this section.

Sensor-based approach

Rivera and Ong (2018) focused on classification of non-manual signals, i.e. grammatical and affective facial expression, of sentences signed by 5 deaf individuals using FSL. The data were collected using Microsoft Kinect v2 sensor which captures the location, movement, and audio using its depth sensor, color camera, infrared (IR) emitter, and microphone array. The data collected is in the form 3D videos. The features that were collected include the face orientations, shape units, and animation units which capture the movements of the eyes, eyebrows, mouth, nose, and head. Classification is performed using Support Vector Machine and Artificial Neural Networks and obtained the highest accuracy of 87% on the emotion classes. Further improvements can be made such as increasing the number of data points used for training and validation. In another paper (Oliva et al., 2018), Kinect sensor is also used to classify FSL signed words. The sensor is used to obtain the location of selected body joints in the Cartesian and Spherical coordinate system. The features are classified using Dynamic Time Warping and Support Vector Machine which obtain a peak accuracy of 95%, recall of 95%, and precision of 95.89%. As of the researcher's review, these papers are the only sensor-based approach to FSL studies, hence a potential for further expansion of the field.

Visual-based approach

Visual-based is the most commonly used approach for performing SLR as it requires no external devices except for a camera that captures image or video feed of sign gestures. This approach mainly utilizes computer vision techniques which obtain useful information from visual inputs. However, image or video data are more susceptible to noise, making it more challenging for classification. Sandjaja and Marcos (2009) maximizes the advantages of sensor-based approach for video classification by capturing signers using a color-coded glove. The authors used 5000 FSL numbers, extracting the fingers position relative to the thumb as their features, and classifying them using Hidden Markov Model. The highest average accuracy obtained in their work is 85.52%. A different approach is performed for classifying 10,000 image-formatted FSL numbers in (Montefalcon et al., 2021). Classification is implemented on the images preprocessed with and without Gaussian Blur. Results show that preprocessing with this technique improves the result. The authors implemented classification using two ResNet models, ResNet-18 and ResNet-50. They concluded hat ResNet-18 leads to over-fitting, having excellent results on the training but not on the validation set. The best obtained accuracy in the paper is 92% and 86.7% on the training and validation set using their fine-tuned ResNet-50 model.

However, normal day-to-day conversations require words or phrases to communicate better. With this, Cabalfin et al. (2012) utilized a Manifold Projection approach in performing classification on 72 common FSL signs. From the training set, reference manifolds are created using the Isomap algorithm. Then, these are used for comparison on the validation set and selection of the closest manifold is done using Dynamic Time Warping and Longest Common Subsequence with 89% as the highest obtained accuracy. A more recent take on this is done by Montefalcon et al. (2023) whose objective was to classify 15 Filipino phrased formatted as video. The author maximized MediaPipe Holistic, an open-source pipeline that captures landmarks coordinates of the human body from images or videos. The paper highlights the use of deep networks, particularly Long-Short Term Memory (LSTM) and ResNet Convolutional Neural Networks. The authors concluded that LSTM performed better, achieving an accuracy of 94%. Using the better performing classifier, feature importance analysis is implemented. Here, isolation of the eyebrows, eyes, and mouth landmarks is done, and the classification performance is checked whether improvements is observed by excluding some or all of these isolated features. Their results show that these features significantly reduced classification accuracy and hence are important features in SLR.

METHODS

This section presents the discussion of the development of the FSLR system using PHCA and other classifiers. Details on data description, data preprocessing and preparation, and the implementation of the PHCA are elaborated.

Data Description

The dataset used in this paper is published by Porton (2023). The dataset consists of images of hands signing letters of the FSL Alphabet. This paper only considers static signs, hence omitting the letters J and Z. Each image is of dimension 300 pixels $\times$ 300 pixels $\times$ 3 channels.

Data Preprocessing and Preparation

Data preprocessing focuses on extracting the important features from the dataset which will be useful for classification. In this section, we discuss the actual preprocessing step and the data preparation process. The framework of the procedures involved in these stages is shown in Fig. 1.

The preprocessing and preparation stage can be summarized into the following parts.

1. **Feature Extraction**: The feature extraction process mainly utilized the MediaPipe Hands pipeline. The pipeline essentially extracts the 3D coordinates of certain landmarks tracked from the hand in the image.

2. **Data Splitting**: Once data preprocessing is performed, data preparation is implemented. The first part of this is to split the data into training and testing. Here, five-fold cross validation is implemented.

3. **Feature Standardization**: The data points are then scaled using the Standard Scaler technique. It follows from the standard normal distribution which makes the mean across data points of each feature vector equal to 0 and scales the data to unit variance.

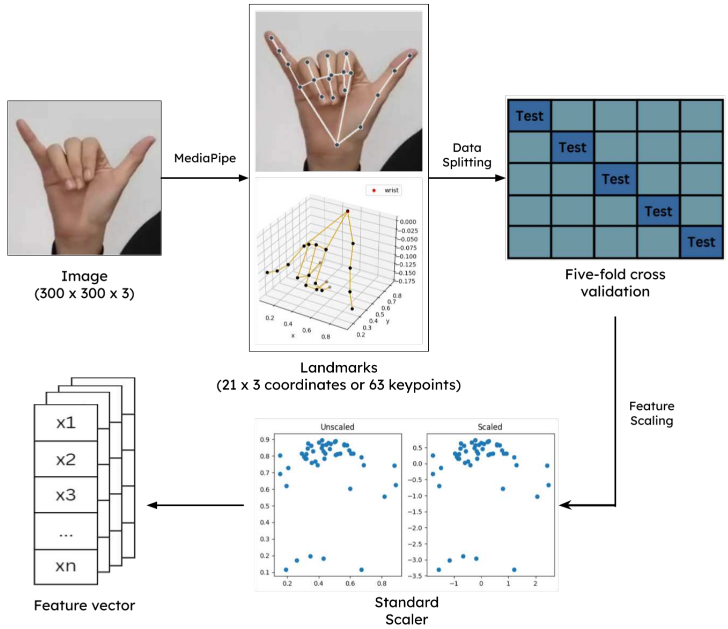

Figure 1. Illustration of the framework of the data preprocessing and data preparation stages.

MediaPipe Hands

MediaPipe Hands is a pipeline whose structure involves a palm detection model and a hand landmarker model. The palm detection model reduces the complexity by estimating first a bounding box on the palm(s). Precise keypoint localization is then implemented using the hand landmarker model to extract the 21 3-dimensional hand-knuckle coordinates. Figure 2 presents all hand landmarks detected by the pipeline. Specifically, each landmark consists of the x and y coordinates and the z-value representing the depth with respect to the camera.

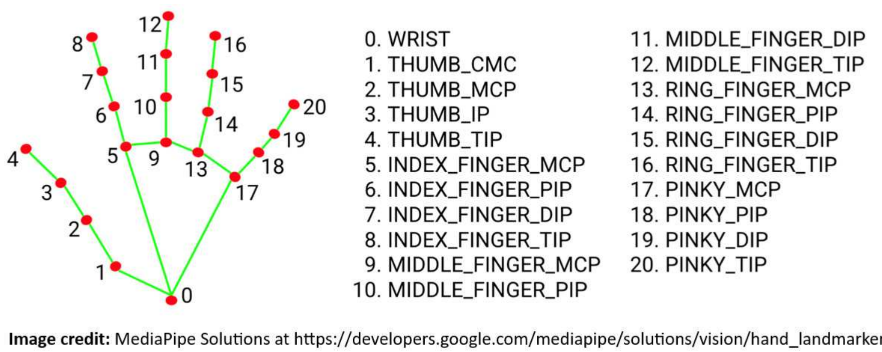

Figure 2. Hand landmarks detected by MediaPipe Hands pipeline.

Five-fold Cross Validation

The five-fold cross validation is a procedure that divides the data into five folds wherein each fold consists of almost the same number of data points. Each validation then allows to split the data into 80:20 train-test ratio. In the first validation, the first fold is considered as the test set while the model is trained using the second to fifth folds. In the second validation, the second fold is considered as the test set and the training set consist of the first and third to fifth folds. This is done until the fifth validation in which the fifth fold is considered as the test set while the first four folds are the training set. This procedure ensures that each data point is considered as part of the training and test sets of the model. The specified performance metrics are then averaged across these five validations.

Classification

The preprocessed and prepared data is now classified using PHCA and other classical classifiers namely LDA, SVM, KNN, CART, and RF. In line with the structure of five-fold cross validation, each run of validation implies a unique training and testing of the models and classification results. Specifically for each validation, the PHCA and other models are trained using the training set consisting of the folds determined by the validation index. Then, the classifiers are validated using the test set also determined by the validation index. The results of this validation are the precision, recall, f1-score, and specificity values for each class considered. The overall accuracy across all is also obtained.

Persistent Homology

Persistent Homology (PH) is a method widely used in Topological Data Analysis (TDA), a growing field of research that incorporates tools and techniques from topology in analyzing data. PH can be used for determining invariant features or topological properties of a space of points that persist across multiple resolutions (Carlsson, 2009; Edelsbrunner and Harer, 2008). These invariant features capture the qualitative properties of data due to their sensitivity to small changes in the input parameters, making PH favored by researchers. PH application extends to different data types such as point clouds, images, time series, etc. In this paper, we focus on the computation of PH on point clouds.

A point cloud (X,d) represents a finite set of points X together with a distance function d. The usual assumption is that X is sampled from an underlying topological space $\mathbb{S}$. However, describing the topology of $\mathbb{S}$ based on sample points of X is not easy. This is where PH can be used.

In computing PH, X undergoes a filtration process, converting the point cloud into a nested sequence of simplicial complexes. A visual description of this process is presented in Fig. 3. This is done by defining a non-negative real number ε that serves as a parameter to thicken X. We denote the thickened point cloud corresponding to the parameter ε as X_ε. As the value of ε increases, simplices are added to the complexes, and a sequence of nested simplicial complexes is formed.

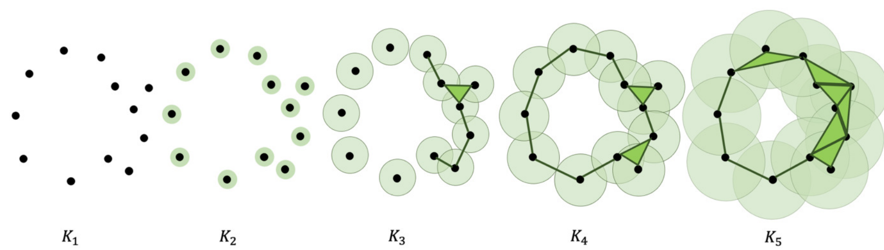

Figure 3. Illustration of a filtration of a point cloud into a nested sequence of simplicial complexes with $K_1 \subseteq K_2 \subseteq \ldots \subseteq K_5$.

Adding new simplices to the complexes can be done using a variety of ways. In this paper, Vietoris-Rips (VR) complex is used. In VR complex, two points x_i and x_j, each of which are initially 0-simplex, are connected when the distance $d(x_i,x_j) \leq 2\varepsilon$. This forms a 1-simplex or a line segment. Adding another point x_k that satisfies $d(x_i,x_k) \leq 2\varepsilon$ and $d(x_j,x_k) \leq 2\varepsilon$ forms a 2-simplex or a triangle. Adding another point x_l that satisfies $d(x_a,x_l) \leq 2\varepsilon$ for $a = i, j, k$ forms a 3-simplex or a tetrahedron, and so on. It is worth noting that the parameter ε is the only parameter changing in this filtration process. The addition of simplices depends on this ε value and the distances of 0-simplices from one another.

In each filtration step, the *homology groups* are extracted. These are invariant features of a topological space that provide important information and can be computed algebraically. The homology of the underlying topological space $\mathbb{S}$ can be approximated by the homology of the simplicial complexes derived from X.

For this, suppose K is a finite simplicial complex and $K_1 \subseteq K_2 \subseteq \ldots \subseteq K_r = K$ is a finite sequence of nested subcomplexes of K. Here, K is called a filtered simplicial complex and the sequence $\{K_1, K_2, \ldots\}$ is the filtration of K. The homology of each of the subcomplex can be computed as follows. For each p, the inclusion maps $K_i \to K_j$ induce $\mathbb{F}_2$-linear maps $\partial_i^j : H_p(K_i) \to H_p(K_j)$ for all $i, j \in 1, 2, \ldots, r$ with $i \leq j$. It follows from functoriality that $\partial_k^j \circ \partial_i^k = \partial_i^j$ for all $i \leq k \leq j$.

Now suppose K_s is a subcomplex in the filtration or a filtered complex at time s. We define the k-th cycle group of K_s as $Z_k^s = Ker\partial_k^s$ and the boundary group of K_s as $B_k^s = Im\partial_{k+1}^s$. Then, the k-th homology group of K_s is given by

$$H_k^s = \frac{Z_k^s}{B_k^s} = \frac{Ker\partial_k^s}{Im\partial_{k+1}^s}$$

Consequently, for $p \in \{0, 1, 2, \ldots\}$, the p-th persistent k-th homology group of K given a subcomplex K_s is

$$H_K^{s,p}(K, K_s) = H_k^{s,p}(K) = \frac{Z_k^s}{B_k^s \cap Z_k^s} = \frac{Ker\partial_k^s}{Im\partial_{k+1}^s \cap Ker\partial_k^s}$$

and the p-th persistent k-th Betti number $\beta_k^{s,p}$ of K_s is the rank of $H_k^{s,p}$.

Simply, for each nonnegative integer p, there exists a p-th homology group $H_p(X_\varepsilon)$ representing X_ε. The 0-th dimensional, 1-dimensional, and 2- dimensional homology groups gives the connected components, holes or tunnels, and voids, respectively. These algebraic structures are homotopy invariant, meaning they do not change when the space undergoes bending, stretching, or other deformations, making them ideal as representation of data.

The result of obtaining the homology of the filtered complexes can be represented using a persistence diagram or barcode. Example of some filtration process and their corresponding diagrams and barcodes are show in Fig. 4. These representations show the appearance (birth) and disappearance (death) of intrinsic topological features, such as homology groups and Betti numbers. In other words, these birth and death values represent the filtration index (or the parameter ε) at which the topological feature appear or disappear, respectively. The lifespan or duration of these topological properties are essential for the qualitative analysis of the topology of the data. Shorter lifespan are often associated with noise while longer ones are the important topological features. This lifespan parameter will be essential for the development of the topology-based classifier PHCA. For a more comprehensive discussion of the computation of PH, the reader is referred to Edelsbrunner and Harer (2008).

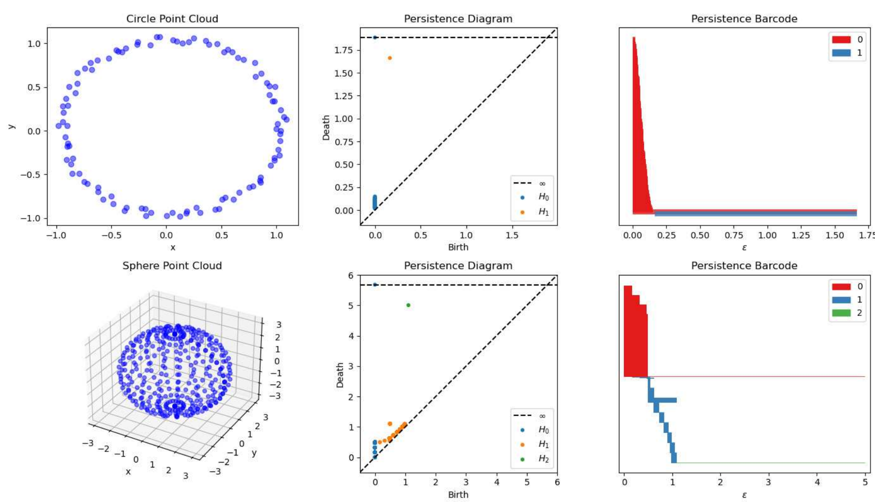

Figure 4. Point clouds and their corresponding persistent diagrams and barcodes obtained using persistent homology with Vietoris Rips filtration. A persistent 1-dimensional hole can be observed for the circle point cloud. Meanwhile, there is a persistent 2-dimensional hole (or void) observed for the sphere point cloud.

Persistent Homology Classification Algorithm

The application of PH in machine learning tasks has been one of the many focuses of many studies in recent years . In (Hensel et al., 2021), a survey is conducted to review and synthesize the current state of the fuse of TDA and machine learning, to which the authors termed as Topological Machine Learning. They divided some of these application into two parts: extrinsic and intrinsic approach.

Extrinsic approach uses topological methods, such as PH, in extracting topological properties which are used as representation of data in the form persistence diagrams. The diagrams are converted into features, using either vector-based or kernel-based representations, which are then fed into the common machine learning models.

On the other hand, the intrinsic approach incorporates TDA in the machine learning model itself. This approach either includes topological information into the design of the model or applies topology to study and improve the model. An example of this intrinsic approach is the developed novel supervised classifier PHCA (De Lara, 2023).

a. **Persistent Homology of a Point Cloud**

Let X be a point cloud. Computing the PH of X implies that the point cloud undergoes filtration and the topological properties of X are recorded and visualized using either a persistence diagram or barcode. Alternatively, this result can be represented using an $n \times 3$ matrix and will be denoted as the persistence $\mathscr{P}(X)$. The number of rows n represents the number of topological features or the total number of 0-dimensional holes, 1-dimensional holes, 2-dimensional holes, and so on, depending on the defined maximum dimensions that can be detected during the filtration. For PHCA, the maximum dimensions *maxdim* used is 0. Meanwhile, the first, second, and third column entries of the q-th row of $\mathscr{P}(X)$ represents the dimension, birth, and death times of the q-th topological feature in the filtration of X, respectively.

Note however that the filtration process cannot be performed in an infinite duration of time. Hence, a maximum scale, denoted as *maxsc*, must be defined. Scale in this context represents the ε value or the distance threshold. In practice,

$$maxsc = \frac{1}{2} max_{x,y}\{d(x,y)\}$$

is used where $d(x,y)$ is the distance of any two points x, y with $x \neq y$ of the point cloud.

b. **Training and Classifying using PHCA**

Suppose X is the training dataset consisting of m-dimensional data points categorized into k distinct classes. More specifically, suppose that $X = X_1 \cup X_2 \cup \ldots \cup X_k$ where each X_i is the set of data points in class i for $i = 1, 2, \ldots, k$. We note that $X_i \cap X_j$ for $i \neq j$ implying that no two classes contains the same data point. Introduced with a new data point α, we want to determine which class does this point belong to.

The training process of PHCA involves computing for the persistence of each class, $\mathscr{P}(X_i)$ for $i = 1, 2, \ldots, k$. Then, the model measures the topological effect of introducing α to each of these classes. For this, the model defines $Y_i = X_i \cup \{\alpha\}$ and compute for $\mathscr{P}(Y_i)$ for $i = 1, 2, \ldots, k$. This process records the changes in PH between X_i and Y_i for $i = 1, 2, \ldots, k$. The new data point is classified to the class which results in the minimum change in PH. This change is measured using the score function discussed in the next section.

c. **Score Function for PHCA**

After training the PHCA model, scoring each of the classes is necessary to choose which class does the new data point α belongs. For this, the model computes for $Score(X_i)$ for $i = 1, 2, \ldots, k$ and compare their results. Recall that the PH of a point cloud can be represented as an $n \times 3$ matrix where n represents the number of topological features and the three columns represent the dimension, birth, and death of each topological feature, respectively. From here, we define the lifespan of the q-th topological feature as

$$l_q = d_q - b_q$$

where b_q and d_q are the birth and death times of the q-th topological feature, respectively. Then, we can define the score function as

$$Score(X_i) = \left| \sum_{q \in \mathscr{P}(Y_i)} l_q - \sum_{q \in \mathscr{P}(X_i)} l_q \right|$$

or the absolute difference of the total sum of lifespan of $\mathscr{P}(Y_i)$ and the total sum of lifespan of $\mathscr{P}(X_i)$. The new data point α is then classified into the class which satisfies

$$\underset{\forall i}{\arg\min}\{Score(X_i)\}$$

Performance Evaluation

To evaluate the performance of the classifiers, five evaluation metrics are obtained. The values of these metrics rely on the confusion matrix corresponding to the predicted classes.

a. **Confusion Matrix**

The confusion matrix is a square matrix representing the true and predicted labels or class of the validation or test set. For each element $a_{i,j}$ of the matrix, this represents the total number of instances that belongs to class i and are predicted to be in class j. From this matrix, the following terms are defined.

- True Positive (TP) - number of instances where the classifier predicts observation under class k to belong to class k.
- True Negative (TN) - number of instances where the classifier predicts observation not under class k to not belong to class k.
- False Positive (FP) - number of instances where the classifier predicts observation not under class k to belong to class k.
- Fale Negative (FN) - number of instances where the classifier predicts observation under class k to not belong to class k.

b. **Classification Report**

From these values, we obtain the five metrics, namely, precision, recall, f1-score, specificity, and accuracy. This comprises of the classification report obtained from the prediction of the model. The first four are obtained for each class while the latter is obtained for the entire test classes. Th description of these metrics are provided in the following:

- Precision describes exactness.

$$precision = \frac{TP}{TP+FP}$$

- Recall or Sensitivity describes completeness.

$$recall = \frac{TP}{TP+FN}$$

- F1-score describes the combination of Precision and Recall. It is defined as the harmonic mean of the two metrics.

$$f1score = \frac{2 \times precision \times recall}{precision + recall}$$

- Specificity describes the ability of the classifier to predict observations not belonging to a class.

$$specificity = \frac{TN}{TN+FP}$$

- Accuracy describes the ratio between the number of correct predictions to the total number of predictions made.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Comparison of Evaluation Metrics

To compare the performance of PHCA with the performance of the other classifiers in terms of the five evaluation metrics, Nemenyi test is implemented. It serves as a post-hoc test for the implementation of Friedman test, a non-parametric equivalent of the repeated-measures ANOVA (Demšar, 2006). The null hypothesis for the Friedman test states that all classifiers are equivalent. If this is rejected, then pairwise comparison of the classifiers is done using Nemenyi test. The performance of two classifiers is significantly different if the corresponding average ranks differ by at least the critical difference

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}$$

where k is the number of classifiers, N is the number of datasets, and q_α are based on the Studentized range of statistic divided by $\sqrt{2}$. The threshold value α used in this paper is 0.05.

Comparison of Predictive Accuracy

In comparing for the accuracy obtained by the classifiers, the authors implemented McNemar's test, a non-parametric test used to analyze statistical difference on the performance of two classifiers. According to Demšar (2006) citing Dietterich, McNemar's test on misclassification matrix is fit for the case where running the classifier multiple times is inappropriate. After the classifier has trained on the train set, a contingency table can be constructed from the predictions made by two classifiers on the test set. The null hypothesis in implementing this test is that the two classifiers have the same error rate. The threshold value α used in this paper is 0.05.

RESULTS AND DISCUSSION

Figure 5 presents the average precision, recall, f1-score, and specificity across the 24 letters or classes obtained by PHCA and the other five classifiers while Fig. 6 shows the corresponding overall accuracy. Here we see that PHCA joined SVM with the 2nd highest value obtained for the average precision, outperformed by RF by approximately 1%. The same results can be observed for the average recall and f1-score for all classifiers. Meanwhile, all models performed well in terms of average specificity, each obtaining a value of 100%. The most accurate classifier for the FSL Alphabet dataset is RF, followed both by PHCA and SVM, then by KNN. The least accurate classifier is CART.

Comparison of Evaluation Metrics

All performance evaluation metric values, including the metric values from each of the 24 classes, obtained by PHCA and the other five classifiers are compiled and compared using Friedman test and Nemenyi test as a post-hoc analysis. Since the resulting p-value of the Friedman test is 1.01×10^{-27} which is less than $\alpha = 0.05$, then there is at least one classifier with a significant different mean of performance evaluation metric from another classifier. Table 1 then presents the Nemenyi table corresponding to the post-hoc analysis implemented on the results. It shows that PHCA, SVM, and RF have a significantly equal mean of performance metrics but significantly different from that of KNN, LDA, and CART. This suggests that PHCA performed similarly with the other two best performing classifiers in terms of all evaluation metrics, RF and SVM. It also shows that these three classifiers performed better than KNN, LDA, and CART.

Table 1. Nemenyi table for the comparison of performance of the classifiers.

	PHCA	KNN	RF	SVM	LDA	CART
PHCA	1					
KNN	0.0014	1				
RF	0.9000	0.0010	1			
SVM	0.9000	0.0011	0.9000	1		
LDA	0.0015	0.9000	0.0010	0.0012	1	
CART	0.0010	0.0010	0.0010	0.0010	0.0010	1

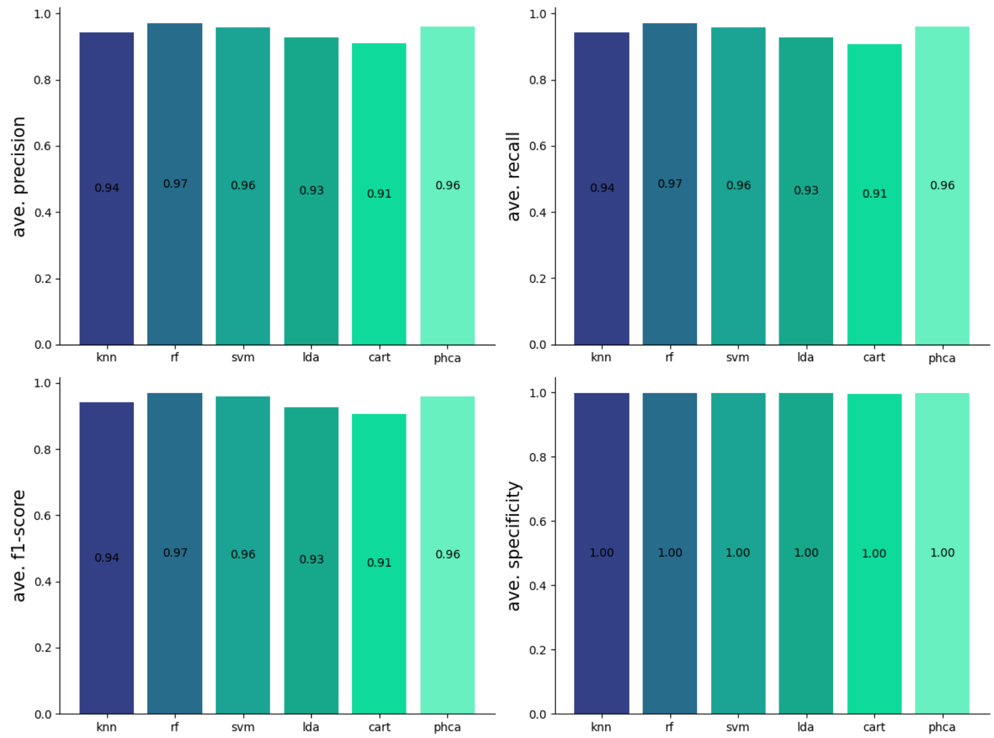

Figure 5. Barplot of the average precision across classes of PHCA and the five other classifiers.

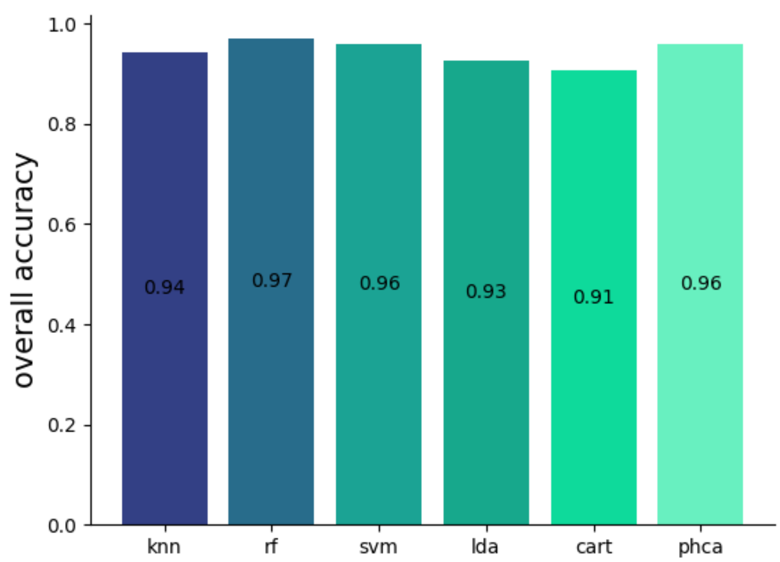

Figure 6. Barplot of the overall accuracy of PHCA and the five other classifiers.

Comparison of Predictive Accuracy

The authors implemented McNemar's test to perform pairwise comparison of the overall accuracy obtained by PHCA with another classifier. The p-values of the tests are tabulated in Table 2. Since the p-value for the PHCA-SVM comparison is approximately 0.614 which is greater than $\alpha = 0.05$, it suggests that the predictive accuracy of the two models have the same error rate, hence similar performance. On the other hand, the PHCA-KNN, PHCA-LDA, and PHCA-CART comparisons show p-values significantly less than $\alpha = 0.05$, implying that there is a significant difference in error rates between PHCA and the other three models. Meanwhile, the PHCA-RF comparison shows a p-value less than $\alpha = 0.05$, indicating significant difference in error rate of the two models. But a closer examine of the contingency table corresponding to this test (shown in Table 3) shows that RF outperformed PHCA, having exclusively making 188 correct predictions while PHCA having only 110, hence the difference in predictive accuracy.

Table 2. McNemar test of the predictive accuracy of PHCA and another classifier.

	KNN	RF	SVM	LDA	CART
p-value	2.72×10^{-16}	6.23×10^{-6}	6.14×10^{-1}	8.87×10^{-25}	9.89×10^{-55}

Table 3. Contingency table obtained from McNemar's test corresponding to the predictive accuracy comparison of PHCA and RF.

		RF	
		Correct	Wrong
PHCA	Correct	7099	110
	Wrong	188	115

CONCLUSIONS

In this study, a novel topology-based classifier called PHCA is utilized to classify images of signed alphabet using FSL. The authors focused on static signs, considering only the 24 letters of the alphabet, excluding letters J and Z. The performance of PHCA in terms of precision, recall, f1-score, specificity, and accuracy is compared with widely used classifiers such as SVM, LDA, KNN, CART, and RF. Results show that in all metrics, the top 3 classifiers are RF, SVM, and PHCA with values significantly similar. Comparison of evaluation metrics obtained by the classifiers is done using Nemenyi test. It shows further that RF, SVM, and PHCA have a significantly equal mean of performance metrics but significantly different from the other three classifiers. This implies that PHCA performed at par with the other two best performing classifiers, RF and SVM, and better than the least performing ones. Comparison of predictive accuracy is also implemented using McNemar's test which reveals the outperformance of PHCA against KNN, LDA, and CART and the similar performance of the model against SVM in terms of accuracy. It also reveals that RF outperformed PHCA in terms of accuracy. Regardless, PHCA have shown excellent performance in terms of classifying images of FSL alphabet with an accuracy of 96%. Further explorations on the applications of PHCA, and TDA in general, to dynamic FSLR could be done. Real-time FSLR system could also be developed by optimizing the computation process of PH.

ACKNOWLEDGMENTS

The authors acknowledge the Institute of Mathematical Sciences and Physics, University of the Philippines Los Baños for allowing the usage of their computer laboratory in the implementation of the methods in this paper.

REFERENCES

Cabalfin, E. P., Martinez, L. B., Guevara, R. C. L., and Naval, P. C. (2012). Filipino sign language recognition using manifold projection learning. In *TENCON 2012 IEEE Region 10 Conference*, pages 1–5.

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308.

Center, P. D. R. and of the Deaf, P. F. (2004). *An Introduction to Filipino Sign Language: (a). Understanding structure*. An Introduction to Filipino Sign Language. Philippine Deaf Research Center.

Cruz, F. and Calimpusan, E. (2018). Status and challenges of the deaf in one city in the philippines: towards the development of support systems and socio-economic opportunities. *Asia Pacific Journal of Multidisciplinary Research*, 6(2):33–47.

Davis, A. C. and Hoffman, H. J. (2019). Hearing loss: rising prevalence and impact. *Bulletin of the World Health Organization*, 97(10):646–646A.

De Lara, M. L. D. (2023). Persistent homology classification algorithm. *PeerJ Computer Science*, 9:e1195.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. 7:1–30.

Edelsbrunner, H. and Harer, J. (2008). Persistent homology—a survey. *Discrete and Computational Geometry - DCG*, 453.

Hensel, F., Moor, M., and Rieck, B. (2021). A survey of topological machine learning methods. *Frontiers in Artificial Intelligence*, 4.

Montefalcon, M. D., Padilla, J., and Rodriguez, R. (2023). *Filipino Sign Language Recognition Using Long Short-Term Memory and Residual Network Architecture*, pages 489–497.

Montefalcon, M. D., Padilla, J. R., and Llabanes Rodriguez, R. (2021). Filipino sign language recognition using deep learning. In *2021 5th International Conference on E-Society, E-Education and E-Technology*, ICSET 2021, page 219–225, New York, NY, USA. Association for Computing Machinery.

Most, T. (2004). The effects of degree and type of hearing loss on children's performance in class. *Deafness and Education International*, 6(3):154–166.

Oliva, K. E., Ortaliz, L. L., Tobias, M. A., and Vea, L. (2018). Filipino sign language recognition for beginners using kinect. In *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, pages 1–6.

Porton, J. G. (2023). Fsl dataset. `https://www.kaggle.com/datasets/japorton/fsl-dataset`.

Rivera, J. P. and Ong, C. (2018). Recognizing non-manual signals in filipino sign language. In *Proc. Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 1–8.

Sandjaja, I. and Marcos, N. (2009). Sign language number recognition. pages 1503 – 1508.

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., and Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking.