2021

Analysis of Sign Language Facial Expressions and Deaf Students' Retention Using Machine Learning and Agent-based Modeling

Marie Alaghband
University of Central Florida

ANALYSIS OF SIGN LANGUAGE FACIAL EXPRESSIONS AND DEAF STUDENTS' RETENTION USING MACHINE LEARNING AND AGENT-BASED MODELING

by

MARIE ALAGHBAND
B.S. Isfahan University of Technology, 2012
M.S. Institute for Management and Planning Studies, 2016

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2021

Major Professor: Ivan Garibay

ABSTRACT

There are currently about 466 million people worldwide who have a hearing disability, and that number is expected to increase to 900 million by 2050. About 15% of adult Americans have hearing disabilities and about every three in 1,000 U.S. children are born with hearing loss in one or both ears. The World Health Organization (WHO) estimates that unaddressed hearing loss poses an annual global cost of $980 billion, including cost of educational support, loss of productivity, and societal costs. These are all evident that people with hearing loss are experiencing several kinds and levels of difficulties. In this dissertation, we are addressing two main challenges of hearing impaired people; sign language recognition and post-secondary education. Both sign language recognition and reliable education systems that properly support the deaf community are essential needs of the globe and in this dissertation we aim to attack these exact problems. For the first part, we introduce novel dataset and methodology using machine learning while for the second part, a novel agent-based model framework is proposed.

Facial expressions are important parts of both gesture and sign language recognition systems. Despite the recent advances in both fields, annotated facial expression datasets in the context of sign language are still scarce resources. In this dissertation, we introduce an annotated sequenced facial expression dataset in the context of sign language, comprising over 3000 facial images extracted from the daily news and weather forecast of the public tv-station PHOENIX. Unlike the majority of currently existing facial expression datasets, FePh provides sequenced semi-blurry facial images with different head poses, orientations, and movements. In addition, in the majority of images, identities are mouthing the words, which makes the data more challenging. To annotate this dataset we consider primary, secondary, and tertiary dyads of seven basic emotions of ”sad”, ”surprise”, ”fear”, ”angry”, ”neutral”, ”disgust”, and ”happy”. We also considered the ”None” class if the image’s facial expression could not be described by any of the emotions. Although we provide FePh as a facial expression dataset of signers in sign language, it has a wider application in gesture recognition and Human Computer Interaction (HCI) systems.

In addition, post-secondary education persistence is the likelihood of a student remaining in post-secondary education. Although statistics show that post-secondary persistence for deaf students has increased recently, there are still many obstacles obstructing students from completing their post-secondary degree goals. Therefore, increasing the persistence rate is crucial to increase education and work goals for deaf students. In this work, we present an agent-based model using NetLogo software for the persistence phenomena of deaf students. We consider four non-cognitive factors: having clear goals, social integration, social skills, and academic experience, which influence the departure decision of deaf students. Progress and results of this work suggest that agent-based modeling approaches promise to give better understanding of what will increase persistence.

This dissertation is dedicated to my mother, who is the light of my life and to my father, who is no longer of this world, but is with me in every moment of this achievement.

I am also dedicating this dissertation to my partner and best friend of all the time, Hamid Reza Maghroor, who loved and supported me unconditionally during all the hardships of the past years.

ACKNOWLEDGMENTS

The research presented in this dissertation would not have been possible without the supervision, persistent help, and advise of my PhD advisor and committee chair Dr. Ivan Garibay. Throughout the past years in the Complex Adaptive Systems Laboratory (CASL), he supported, advised, and guided me in all situations and believed in me like a father. I am forever thankful for his support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
ASL	American Sign Language
BM	Boltzmann Machine
CNN	Convolutional Neural Network
DBM	Deep Boltzmann Machines
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
EM	Expectation Maximization
GAN	Generative Adversarial Network
HCI	Human Computer Interaction
HMM	Hidden Markov Model
KNN	K-Neareset Neighbors
ML	Machine Learning
RBM	Restricted Boltzmann Machine
RNN	Recurrent Neural Network
SPN	Sum-Product Network
WER	Word Error Rate

CHAPTER 1: INTRODUCTION

Language is an important key in human lives. It is a complex system that helps us not only to effectively communicate with others, but also express our personalities. We utilize words, gestures, and tones of voice in a multitude of situations to communicate with others and express our feelings, desires, and queries to the world around us.

Sign language is the natural language of people with severe or profound hearing loss. Over 5% of the world's population (about 466 million people worldwide) have disabled hearing, and by 2050 this number will increase to 900 million people–one in every ten (based on the most recent report of World health Organization (WHO) [2]).

Sign language was not truly defined until the first half of 20th century. It was thought that it was a substitute for speech and a way to express natural language words with signs. Later on, in the second half of the century, it was understood that signs performed in sign language do not stand for natural (spoken) language words. In fact, signs performed in sign language express meanings and not spoken language words. Sign language is not universal and is unique to each country or region [3]. There are over 135 different sign languages all around the world. Countries that share a spoken language do not usually share the same sign language. For example, United States of America, England, and Australia share English as a spoken language, but their sign languages are different [4].

It is also noteworthy that sign languages have their own grammar, syntax, and structure. The grammar used in sign language is completely different from the spoken language. For example, a single sign performed in sign language can be translated to a sentence in spoken language. In addition, similar to all languages, sign languages of each region or country grow and change over time [5].

Sign language, as a well defined gesture inventory, is composed mainly of hand gestures and facial expressions (usually known as manual and non-manual signals, respectively). Manual signals are performed using hand shape, orientation, location, and motion. These signals are widely used to convey meanings. On the other hand, other body parts, such as eye gaze and movement, lip pattern, mouthing, and head orientations [6, 7, 8, 9], are usually utilized to perform non-manual signals. These signals are usually used to convey grammatical and emotional information.

Sign language users typically utilize manual signals to convey information such as words or sentences. However, sometimes a single manual signal (hand gesture) may convey different meanings. In this case, the signer performs manual signals combined with non-manual signals (i.e., signer performs facial expressions during the communication) to deliver the exact meaning and sense to the signed gesture. In other words, non-manual signals give auxiliary information to convey the correct meaning and eliminate misunderstandings and mistranslations. Therefore, the combination of both signals creates useful meaning in sign language that makes it unique and complex [10]. Both signals are complementary to each other, and one is incomplete without the other [11].

1.1 Sign Language Recognition Systems

Due to the importance of sign language among the hearing impaired and deaf community, it has been vastly studied by researchers in various fields including computer science, linguistics, computer engineering, and education. Among these different aspects of sign language study, sign language recognition is a research area with the objective of building methods and algorithms to identify signs and perceive their meanings [12].

Over the past number of years, the majority of the works presented in the sign language and gesture recognition literature have focused on recognizing either manual or non-manual signals. More specifically, only hand gestures and shapes have been considered to address the problem of sign language recognition [13]. As previously mentioned, despite the importance of manual signals, non-manual signals' role in expressing grammar and emotion is undeniable [11]. Therefore, both categories of signals are important for conveying information and communicating with others. Sign language recognition systems without considering facial expressions are thus incomplete [6].

Research in sign language recognition systems can be categorized in two main groups: vision-based (without using any special hardware) and hardware-based recognition systems. The hardware-based recognition systems use datasets that are collected utilizing special colored gloves [14, 15], special sensors, and/or depth cameras (such as Microsoft Kinect and Leap Motion) to capture special features of signers' gestures [16, 17, 18, 19]. On the other hand, some researchers argue that although utilizing these tools ease the process of capturing special features, they limit applicability where such hardware are not available. Therefore, they propose vision-based sign language recognition systems utilizing datasets collected by regular cameras [20, 21, 22, 23, 24]. There are very few researchers that consider both types of data [25].

1.1.1 Vision-Based Sign Language Recognition Systems

Vision-based sign language recognition systems are developed using computer vision Artificial Intelligence (AI) techniques such as machine learning (ML) and deep learning (DL). Though research in computer vision sign language and gesture recognition is challenging, it enabled the creation of a real time interpreter system to solve communication barriers between deaf and hearing people who do not understand sign language [26].

AI is a broad field of science and research that has been applied in a vast variety of research areas such as computer vision, computer science, robotics, gaming, healthcare, etc. AI is usually referred to as the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent creatures such as human beings. In another word, by the definition of Professor John McCarthy, the goal of AI is to develop machines that behave as if they were intelligent [27].

In order to apply AI techniques that enables computers to mimic human intelligence, scientists use different kinds of tools. While some of these tools may seem disjointed (i.e., logic, if-then rules, decision trees, etc.), some may be a subset of one another such as machine and deep learning. For example, deep learning is a subset of machine learning, while machine learning is a subset of AI. Neither machine learning nor deep learning are brand new ideas, but the way they are used is changing and improving constantly. In the following subsections, both methods are further discussed.

1.1.1.1 Machine Learning (ML)

As a subset of AI, machine learning goes back to 1950 when the concept of a "learning machine" was discussed for the first time [28].

Machine learning includes complicated statistical and data science techniques that enhance a machine's ability to learn and experience by itself, i.e., without outside help from human beings. Machine learning systems usually have an objective function and are used to minimize an error or maximize their correct prediction likelihood. In short, this method is an optimization algorithm that can minimizes its errors by learning. In the machine learning method, there are several layers of data, consisting of some number of hidden layers (i.e., layers that are not in the input or output layer). Despite many successful achievements of machine learning methods, experimental results illustrate that these techniques have increased ability as the hidden layers of a learning process increase. In this case, deep neural networks are used, i.e., where the hidden layers are 3 or more [29]. Today, machine learning algorithms, deep or not, are broadly used in supervised learning, unsupervised learning, and reinforced learning.

As mentioned before, modeling and analyzing complex real-world data is complicated and can be approached by different methods. From the machine learning perspective, one way is capturing relevant features by developing robust features and feature extractors. This method requires careful engineering and expertise of the data as well as domain-specific features, which is expensive and time-consuming [30]. After a year of experience, [31] discusses that due to these problems, conventional machine learning techniques showed limited ability in processing natural data in their raw format.

The alternative is using unsupervised learning [32, 33, 34]. While supervised learning uses labeled data to train the machine, unsupervised learning generates answers on unlabeled data. Unlabeled data is plentiful, easy to obtain, and does not require hand-engineered features. Another advantage of unsupervised learning is that more layers of representation can be extracted, and these layers can be stacked to create deep networks [30].

1.1.1.2 Deep Learning (DL)

Deep learning is a class of machine learning techniques that tries to mimic the human brain's neurons. Although deep learning is a subset of machine learning, it refers to a wider class of machine learning techniques and architectures. Some of the more common techniques and architectures that will be addressed in this section include Recurrent neural network (RNNs), Convolutional neural networks (CNNs), Deep neural networks (DNNs), Deep belief networks (DBNs), Boltzmann machines (BMs), Restricted Boltzmann machines (RBMs), Deep Boltzmann machines (DBMs), Deep auto-encoders, Hidden Markov models (HMMs), Generative Adversarial Nets (GAN), and Sum-Product networks (SPNs). Depending on the way these architectures and techniques are used, most of the works can be categorized in three main classes [35, 36]:

- Generative or Unsupervised deep learning architectures

- Discriminative or Supervised deep learning architectures

- Hybrid Generative-Discriminative deep architectures

The traditional way of categorizing deep learning architectures did not include the third category (i.e., hybrid deep architectures) and classified the architectures only in the first two classes: deep generative models (e.g., RBMs, DBNs, DBMs and Regularized auto-encoders) and deep discriminative models (e.g., DNNs, RNNs and CNNs). This two-way classification without the third category missed out on a key benefit gained in deep learning research: generative models can greatly improve the training of DNNs and other deep discriminative models via improved regularization or optimization [36, 35]. Below, further descriptions of each of these main classes of deep leaning networks will be illustrated.

Generative or unsupervised deep learning architectures. Whenever no information about target class labels is available, generative, or unsupervised learning architectures will be used in order to capture high-order correlations of the observed or visible data for pattern analysis. A key concept of applying generative architectures for pattern recognition is pre-training. Pre-training arises from learning all the layers of deep networks, starting from the lower layers without relying on all the layers above and continuing this bottom-up process in a layer-by-layer manner. This makes the model more efficient.

Unsupervised learning refers to a learning algorithm that generates answers on unlabeled data. Deep generative or unsupervised learning models and techniques are usually easier to interpret and compose, and are better at handling uncertainty and embedding domain knowledge. Several examples of this class include RBMs, DBNs, DBMs, HMMs, auto-encoders, SPNs and RNNs. Energy-based deep models that include auto-encoders are the most common among the various sub-classes of generative deep architectures. The original form of deep auto-encoders as well as most other forms of deep auto-encoders (e.g., transforming deep auto-encoders, predictive sparse coders and their stacked version and de-noising auto-encoders and their stacked versions) are in the generative or unsupervised deep architecture class.

While a detailed discussion of these techniques, architectures, and learning algorithms is beyond the scope of this section, an example of generative or unsupervised deep learning architectures called deep Boltzmann machine (DBM) is discussed. DBM is a new learning algorithm for the general Boltzmann machines that contains many layers of hidden variables [37]. In addition, in DBMs, there is no connection between variables within the same layer. DBMs use two quite different techniques for estimating the two types of expectations: a variational approximation that tends to focus on a single mode for the data dependent expectations, and persistent Markov chains for data-independent expectations. These techniques make it practical to learn Boltzmann techniques with multiple hidden layers and millions of parameters. As mentioned before, using a pre-training phase can make this model more efficient. Salakhutdinov and Hinton show that the DBM better learns generative models if the pre-training phase is included. Moreover, they also show that it performs well on handwritten digit and visual object recognition tasks.

Discriminative or supervised deep learning architectures. Supervised learning is the most common form of machine learning [31] in which both input and desired output data are provided. The input and output data are used for classification and are labelled to provide a learning basis for future data processing. In other words, the machine in supervised learning is trained with labeled data. It is shown an image and outputs a vector of scores, which is compared with other scores for each category of data. The difference between the output score and desired pattern scores is measured with the help of an objective function to find the level of error. To decrease error, the machine modifies its internal adjustable parameters, often called weights. Deep learning systems are usually trained with hundreds of millions of these adjustable weights and large amounts of labelled examples (hundreds of millions). Overall, discriminative or supervised learning takes labelled datasets (paired input subjects and desired output) and learns to label new datasets by extracting and gaining information from it.

Deep discriminative or supervised learning models and techniques have some benefits over the unsupervised approach, but also have limitations. To number some of the benefits, they usually are more efficient to train and test, more flexible to construct, and are more suitable for end-to-end learning of complex systems.

One example of discriminative or supervised deep architectures is convolutional neural networks (CNN). CNNs are a family of multi-layer neural networks particularly designed to process data in the form of multiple arrays: 1D for signals and sequences, 2D for images or audios, and 3D for video or volumetric images [31]. CNNs are the first truly successful deep learning approach [38].

CNN is highly effective in Computer Vision, Computer Recognition, Image Classification, Object Detection and many other topics related to Image Processing. Thus, images are the most common input for the input layer. In images, neurons are arranged in a three-dimensional structure including "Width," "Height," and "Depth," which in RGB images the Depth is equal to three. It should be noted that each pixel of the raw input image will be considered as a neuron in the CNN process.

CNNs have three main layers including the Input Layer, Feature Extraction or Learning layer, and Classification Layer. In addition, the feature extraction layer has two main layers: convolution, and pooling. These two layers repeat the pattern of the algorithm sequentially. The convolution layer is the heart of the CNN's architecture and shares the weight of the neurons for the input data. In other words, the convolution layer defines a mathematical operation to determine the rule for neuron weight. The major components in convolution layers are Filters, Activation maps, Parameter sharing, and Layer-specific hyper-parameters. The pooling layer sub-samples the data, which is generated from underlying convolution layers, and then gives this new set of data to the convolution layer in the higher level of CNN system. Finally, the Classification Layer will classify the input data to one class. In the CNN process, all layers are connected to all the neurons of the previous layers.

Hybrid generative-discriminative deep architectures. This class of deep learning neural networks takes advantage of both generative and discriminative deep architectures. The final goal of the hybrid deep architecture is exploiting a generative component to help with discrimination. This goal can be achieved in two ways: first, discrimination which is assisted with the outcomes of generative networks, and second, the use of discriminative criteria to estimate the parameters in any of the deep generative networks. Generative modeling can help with discrimination from two viewpoints: the optimization and the regularization perspective [34].

Some examples of this class of deep architectures are hybrid DBN, GAN, DNN-CRF, hybrid deep architecture of DNN and HMM, generative DBN used to initialize the DNN weights, deep generative model of DBN with the gated MRF, and generative models of DBN used to pre-train deep convolutional neural networks (deep DNN).

1.1.2 Multi-Modal Continuous Sign Language Recognition System

The aim of this dissertation is to enhance and propel sign language recognition systems by considering more that one signal (i.e., two signals of hand shapes and facial expressions) in sign language recognition frameworks.

We started to achieve this goal by introducing the first real-life annotated sequenced facial expression dataset in the context of sign language. This dataset not only provides an annotated facial expression dataset with different head poses, orientations, and movements, but also contributes a sign language dataset with both hand shape and facial expression labels with attributions in multimodal works to the field.

1.2 Student Retention Models

Post-secondary persistence refers to the likelihood of student retention in post-secondary education (e.g., university, collage), especially after the first year of enrollment. Since retention in post-secondary education affects college students in many different aspects [39], it has received considerable attention in the last five decades [40]. Students' retention, program completion, and graduation, advances the overall quality of life for people with and without disabilities [41]. Students with disabilities may have sensory, mobility, mental, emotional or cognitive disabilities. Because of these disabilities, disabled students often encounter more barriers than other students and they complete post-secondary education at lower rates [42].

Deafness or severe hearing impairment is considered as a kind of sensory impairment and a disability [43]. Compared to the general student population, deaf students find the transition to post-secondary setting more problematic [44, 45]. Based on the National Deaf Center's (NDC) most recent report, about 1.3% of all currently enrolled college students are deaf [46]. Although post-secondary enrollment rates for deaf people have increased since the 1980s, the completion degree college rate is still fewer than their hearing peers [47]. These statistics show that there are many deaf students who face obstacles preventing them from completing their post-secondary degree goals. Therefore, increasing the persistence rate in these students plays an integral role in increasing education and work goals of deaf students.

1.3 Statement of Contributions

This dissertation aims to continue having contributions in the field of sign language recognition research. Primarily, the Facial Expression Phoenix (FePh) dataset provides the first annotated vision-based publicly available sequenced facial expression dataset in the context of sign language. The introduction of the FePh dataset has multiple contributions:

- Providing a sequenced facial expression dataset to the fields of facial expression, sign language, and gesture recognition.
- Attributing highly used hand shapes with their associated performed facial expressions.
- Illustrating the relationships between hand shapes and facial expressions in sign language.
- In conjunction with RWTH-PHOENIX-Weather 2014 and RWTH-PHOENIX-Weather 2014 MS Handshapes datasets, constitute the first sign language data with both handshapes and facial expression labels.

With the introduction of the FePh dataset, we are capable to propose and introduce novel vision-based sequenced multi-modal sign language recognition frameworks. Since both hand gestures/shapes and facial expressions are integral parts of sign language and there is a distinguished correlation between hand shapes and facial expressions, multi-modal sign language recognition algorithms that extract features from both modals of hand shapes and facial expressions are going to enhance the accuracy and validity of the architectures. In addition, it will propel research in facial expression, multi-modal sign language, and gesture recognition fields of research.

In addition, to the best of our knowledge, the proposed agent-based model for the problem of post-secondary persistence of deaf students is the first agent-based modeling simulation for measuring the influence of non-cognitive factor on post-secondary persistence in deaf students. All previously published research in the field of deaf students' persistence were theoretical works and have never addressed the problem with an either mathematical or agent-based modeling approach.

1.4 Statement of Originality

Parts of this work have been included in conferences and preprints publicly available. Other than the works presented and discussed in the manuscripts that follows, the rest of this dissertation has not been published publicly at the time of writing.

- Alaghband, M., & Garibay, I. (in-press). Effects of Non-Cognitive Factors on Post-Secondary Persistence of Deaf Students: An Agent-Based Modeling Approach. 2020 Conference of the Computational Social Science Society of the Americas (CSS 2020).
- Alaghband, M. , Yousefi, N. , Garibay, I.. "Facial Expression Phoenix (FePh): An Annotated Sequenced Dataset for Facial and Emotion-Specified Expressions in Sign Language". World Academy of Science, Engineering and Technology, Open Science Index 171, International Journal of Electronics and Communication Engineering (2021), 15(3), 131 - 138.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews sign language literature which generally can be divided in two main categories of sign language recognition and translation. Figure 2.1 shows a schematic diagram of studies and research conducted in sign language literature.

Following the two main categories in sign language literature, this chapter is carried out in two main sections. Section 2.1 focuses on studies addressing sign language recognition, which typically employ sign language recognition through hand gestures recognition, facial expressions recognition, or combined hand gesture and facial expression recognition. Section 2.1.4, on the other hand, focuses on the studies addressing sign language translation in a machine translation context.

2.1 Sign Language and Facial Expression Recognition Literature

As shown in the sign language recognition branch of Figure 2.1, works done in the field of sign language recognition can be categorized in three subcategories: hand gesture, facial expression, and combined recognition. These three categories are described below in subsections 2.1.1, 2.1.2, and 2.1.3, respectively.

2.1.1 Hand Gesture Recognition

An efficient hand gesture recognition system can be an integral potential part of many applications such as natural human-machine interaction (HCI) [48], Virtual object manipulation, Interaction with multimedia and games, Smart houses, Infotainment systems in vehicles [49], and sign lan-

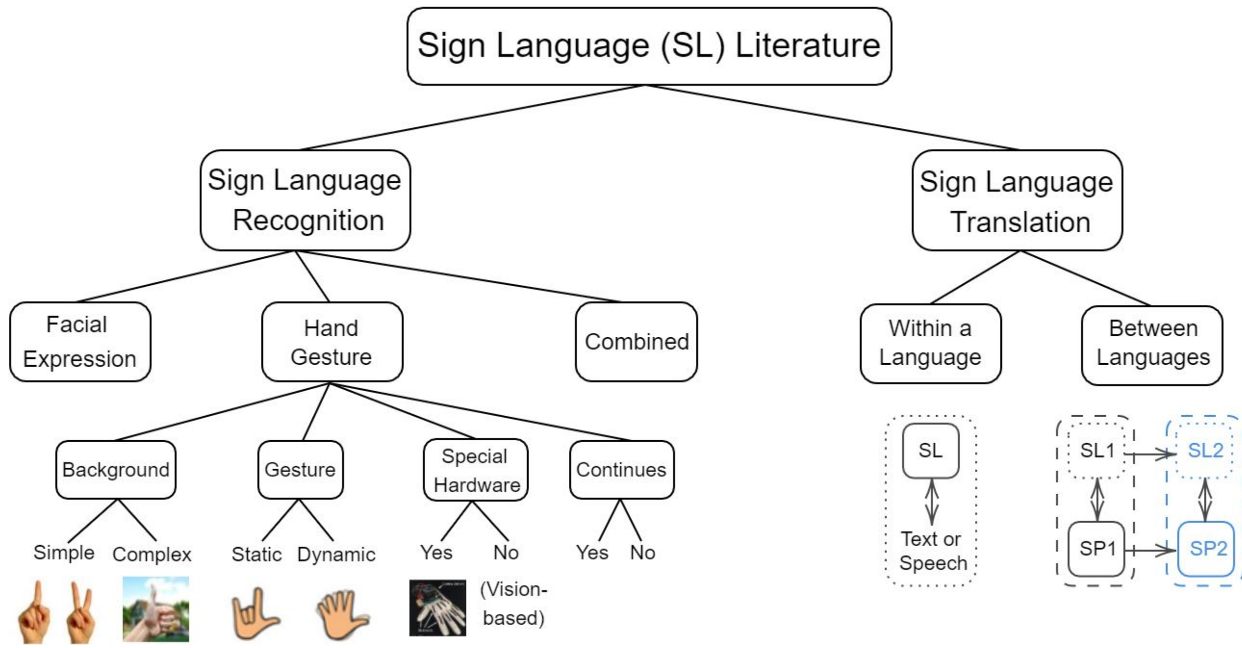


Figure 2.1: Schematic diagram of conducted studies in sign language literature. As the diagram shows, sign language literature in the field of engineering and computer science can be categorized in two main branches of sign language recognition and translation. The problem of sign language recognition by itself can be divided in three branches of facial expression, hand gesture recognition, and combined or multi-modal sign language recognition.

guage recognition [50].

Hand gestures are an important part of sign language that can be described as a combination of hand movement, shape, orientation, alignment, and position of the fingers to the hands and body[51]. Some signs involve hand movements and motions in a large area, while some others involve one finger only [52].

In other words, high inter-class similarities, large intra-class variations, and constant occlusions in hand shapes creates vast variation and complexity in sign language [53] and makes sign language recognition a challenging task. Due to this complexity, hand gesture recognition problems can be approached from different aspects. As Figure 2.1 illustrates, backgrounds, gestures, special hardware usage, and continuity are the four main parts of the hand gesture recognition problem. Any different combination of these parts can be used while addressing the problem of hand gesture sign language recognition. However, we divide these studies in two main categories based on the usage of any kind of special hardware. In other words, studies are categorized in two categories of vision-based (without the usage of any special hardware) and hardware-based (with the usage of special hardware such as dept cameras or special hand gloves) hand gesture sign language recognition. Below, some studies that address the problem of hand gesture sign language recognition in hardware-based and vision-based categories are further discussed.

Hardware-based studies. A recent study done by [53] addresses the static alphabets and numbers sign recognition in American sign language (ASL). A novel method based on multiview augmentation and inference fusion using depth images of two public data sets of ASL static signs captured by Microsoft Kinect is introduced. In their method, they first retrieve 3D information of depth images and then generate more data from different perspectives. Therefore, they simulate realistic perspectives of sign gestures and comprehend information from each of these perspectives. In the next step, the final prediction of each gesture recognition is shown as output. Experimental results based on the ASL benchmark data set of [17] shows an accuracy rate of 93% to 100%, and on the NTU digit data set of [18] the accuracy rate was shown to be 100%.

The study of [54] contributes an automatic recognition of 24 static alphabets and numbers from 0 to 9 in ASL. To take advantage of the depth images, the Kinect sensors are used to capture signs. The proposed recognition system uses an CNN classifier and it is shown that the although the method achieved an accuracy rate of 94.6774%, the accuracy rate can be improved as the number of training data images from different subjects increases. In the same year, [55] uses Leap motion depth sensors to recognize ArSL static and dynamic gestures. The studied dataset consists 28 alphabets, the first 10 numbers, and words including nouns and verbs in ArSL. In addition, a segmentation method to segment a sequence of continuous signs is addressed. Some well-known machine learning algorithms such as SVM, KNN, ANN, and DTW methods are used for implementation. Two main feature sets of palm and bone are considered for the proposed models. Several experiments were performed for the proposed models and it is concluded that KNN outperforms other methods in recognizing both palm and bone features set while achieving an accuracy rate of 99% and 98%, respectively. On the other hand, DTW dominates other methods using the same feature sets with the accuracy rate of 97.4% and 96.4%, respectively.

While using the Microsoft Kinect sensor, [56] develops a novel 3D CNN framework to extract discriminative spatial-temporal features from a raw video stream. With the advantage of using a Microsoft Kinect sensor, multi-channel video streams including color image, depth map, and body joints location were recorded simultaneously. Comparing their proposed method (3D CNN) with the baseline method (Gaussian Mixture Model-Hidden Markov Model- GMM-HMM) with the average accuracy rate of 90.8%, it was shown that the 3D CNN is capable of increasing the accuracy of the method while using the multi-channels and was 94.2% accurate.

In 2017, [57] presents a multi-sensor framework to recognize Indian sign language words using both Leap motion and Kinect sensors. By using these two depth sensors, the authors took advantage of inputting sign language gestures from two different angles which helps the framework record the gestures efficiently and completely. A dataset containing 25 dynamic isolated Indian sign language words signed by 10 different signers 8 times each for each sign is collected. While applying the proposed method using Coupled HMM (CHMM), the model achieved 90.80% accuracy. In a similar study using a multi-model framework to capture finger and palm positions, [58] recognized 50 words signed gestures independently using a Leap motion and Kinect dataset. To improve the accuracy of the proposed system, the collected data from both Leap motion and Kinect was combined and it was shown that the accuracy of the system improves to 97.85% for single and 94.55% for double hand gestures. HMM and bidirectional Long Short-term Memory neural network (BLSTM-NN) are used for the recognition of single and double hand gestures, respectively.

In 2014, [59] proposes a systematic feature selection method that in the context of sign language recognition was tested on the first 10 numbers of American sign language numbers (i.e., 0 to 9). To illustrate the computational results of sign language recognition, the images of the first 10 numbers of ASL at the Massey University data set for hand gestures proposed by [60], was used. They showed that the proposed feature method selection, which minimizes the feature vector while maximizing the F1 score for the classification system, is able to achieve an accuracy rate of 97.7%. With the goal of demonstrating a novel combination of feature and optimization technique, [49] proposed a combination method of the ANNs and Genetic algorithm, named the Non-dominated Sorting Genetic Algorithm II (NSGA-II). The data set of Massey University by [60] was used to recognize 36 hand gestures (26 letters and 10 numbers of American sign language). With the proposed method, authors achieved up to an 98.61 accuracy rate. In addition, they compared several commonly used feature extraction methods on the same data set in terms of accuracy and computational cost. Using the same data set, [61] dealt with using CNN for recognizing sign language hand gestures. They considered both static and dynamic American alphabets and numbers (36 gestures in total) and showed that the proposed CNN model achieved an accuracy rate of 96%.

Due to some of the advantages of glove-based recognition methods, such as being independent of locomotion during the recognition process and environmental lightening conditions they are widely used for sign language recognition. In the context of using a data glove for recognizing the sign language, several studies can be reviewed.

[62] introduces SIGMA that uses both a data glove housing a 6-DOF IMU and nine flex sensors for both static and dynamic hand gesture recognition. Static and dynamic gestures are not considered in a continues gesture recognition format. In other words, the signer must stop to punctuate the ends of each gesture. With this, four states for the static gesture, and five, six, and seven states

for dynamic gestures with same hand form, changing hand forms, and multiple movements are considered, respectively. This is one of the main limitations of the proposed method, SIGMA. In total, 30 data samples are used for each of the gestures with 17 features. By developing a Hidden Markov Model (HMM), the authors achieved an accuracy rate of 79.44%. In another study by [63], the authors recognized single handed Indian sign language alphabets (8 letters), American sign language alphabets (A to Z), and sign numbers (0 to 9) based on hand kinematics while using a data glove. After recognizing alphabets and numbers, they were then translated into speech with the usage of a label matching and speech data base. The proposed method achieved an accuracy rate of 96.7%.

Vision-based studies. Although utilizing special hardware such as data gloves or depth cameras eases the process of capturing special features, they limit applicability where such hardwares are not available. Therefore, vision-based sign language recognition systems utilizing datasets collected by regular cameras are proposed [3, 20, 22, 23, 64]. In vision-based studies, interaction between humans and computers for gesture recognition is involved.

One recent work done by [65] addresses real-time sign language recognition using vision-based machine learning and deep learning methods CNN and RNN. They run a deep CNN for spatial features, and an RNN for temporal features. With a data set including 46 gestures of words in Argentinean sign language (LSA), the proposed method achieved an accuracy rate of 95.2%.

A study conducted by [66] proposes a novel end-to-end embedding of a CNN into an HMM while treating the outputs of the CNN as true Bayesian posteriors. Using 3 publicly available benchmarks in continuous sign language datasets (RWTH-PHOENIX-Weather 2012, RWTH-PHOENIX-Weather Multisigner 2014, and SIGNUM single signer), computational results achieve about 15% lower Word Error Rate (WER). On the first two datasets (RWTH-PHOENIX-Weather 2012 and SIGNUM), the best known WER is achieved:30% and 7.4%, respectively. On the third dataset (RWTHPHOENIX-

Weather 2014 Multisigner), which has more than 1000 vocabularies, a lower WER of 38.8% is achieved. In addition, the authors released a one-million-hand image dataset for public use, which is a valuable benchmark for evaluating state-of-the-art methods.

One year later, [67] presented a new network called SubUNets to improve the results of [66]. SubUNets are a series of specialized expert systems that are introduced to solve sequence-to-sequence learning problems while allowing authors to use domain–specific expert knowledge in the system. In order to illustrate the computational results, SubUNets are evaluated using the one million-hand image dataset, which shows an improvement of about 30% in frame-level accuracy. In the same year, [68] proposed a weakly supervised deep architecture with recurrent convolutional neural network for the continuous sign language recognition. Training their model on the same benchmark dataset of the RWTH-PHOENIX-Weather multi-signer 2014, the authors observe that in various settings, the (WER) of the proposed model achieved comparable results to state-of-the-art models without using extra supervision.

In another study by [69], an iterative re-alignment approach applicable to visual sequence labelling tasks is presented. To model the algorithm, a deep hybrid CNN-BLSTM network was embedded into a Hidden Markov Model (HMM). Results on two publicly available datasets featuring over 1000 classes outperform the state-of-the-art by about 10% absolute WER. In a more recent study by [70] and similar to the [68], a recurrent convolutional neural network for a continuous sign language recognition system is proposed. In contrast to the previous state-of-the-art models, the recurrent neural network in their work is used as the sequence learning module of the framework. By evaluating the proposed method on two publicly available datasets of RWTH-PHOENIX-Weather multi-signer 2014 and SIGNUM, the representation and performance of the model is improved.

In contrast with the majority of vision-based research that considers one gesture (static vs. dynamic), [71]'s study considers both static and dynamic letters in ASL and input webcam image data is stored in the database. A SVM-HMM method is proposed for hybrid arrangements that can perceive single hand motions. Performance metrics of the system are accuracy, sensitivity, precision, FNR, and FDR. The results demonstrate constant improvement of proficiency and robustness via the proposed system.

2.1.2 Facial Expression Recognition

As previously mentioned, sign language is a combination of hand movements and facial expressions. Facial expressions such as eye gaze recognition or direction, eyebrows, eye blinks, and mouth plays an integral role in conveying emotions, feelings, and/or grammar [72]. Signers use facial expressions to support grammatical constructions, called grammatical Facial Expressions (GFEs). GFEs are considered in morphological and syntactic levels of sign language and help in eliminating the ambiguity of signs [16]. Therefore, sign language recognition systems without the facial expression recognition that expresses grammar and emotion are incomplete [7].

For the first time, [73] discussed the importance of facial expressions and addressed it as one of the most powerful and natural signals expressed universally to convey human emotions and intentions. Due to the practical importance in several different fields of research such as sociable robotics, driver fatigue surveillance, sign language recognition, and many other HCI systems, numerous research conducted facial expression recognition studies[74].

In sign language recognition systems, facial expressions play an important role in conveying grammatical and emotional information. Conducted research in this field can be categorized in two main groups depending on the special hardware usage: vision-based (without using any special hardware) and hardware-based recognition systems. The hardware-based recognition systems usually use special sensors such as Microsoft Kinect or Leap Motion to capture special features of signers' gestures. [16] uses Microsoft Kinect to present a new model for automatically recognizing

Grammatical Facial Expressions (GFEs) at the discourse syntactic level.

Since deep neural networks outperform traditional models in numerous fields of computer vision, much research introduces novel vision-based deep learning models to capture and recognize facial expressions. To address a few, [75] propose a novel multi-region ensemble CNN to capture both global and local features from multiple human face sub-regions. [76] presents a new hybrid CNN-Recurrent Neural Network (CNN-RNN) method for facial expression recognition in images of two datasets, the MMI facial expression and the Japanese Female Facial Expression (JAFFE) Database. [77] demonstrates a novel Deep CNN method to learn from noisy labels, using facial expression recognition as an example. In 2015, [78] presented a joint fine-tuning framework using deep CNNs to address the facial expression recognition problem. The proposed framework has two different models: one for capturing temporal appearance features from image sequences and one for temporal geometry features extracted from temporal facial landmark points. For further study on facial expression recognition methods, please see [79], [74], and [80] surveys.

2.1.3 Combined hand gesture and facial expression recognition

As previously discussed, for complete sign language recognition systems, both manual and non-manual signals must be integrated. Although most research studies each category of signals separately, little research proposes combined hand gesture and facial expression frameworks. Frameworks that utilize multi-signals, considering signals as different parts of the human body for the gesture or sign language recognition systems, are called multimodal methods (some of the works in the literature may also call these methods as multi-semantics such as [81]). Here, we review some of most recent and related multimodal studies.

The study conducted by [11] presents a multimodal HMM-based system to recognize hand gestures as manual signals and head movements as non-manual signals simultaneously. Later on, [8] proposed a new multimodal framework, using Hierarchical Conditional Field (H-CRF) and Support Vector Machine (SVM) to recognize hand gestures and facial expressions, respectively. Two years later, the authors proposed another multimodal framework using three cameras capturing three different directions [82]. Another study by [83] presents a novel multimodal framework combining both face and body gestures using Histogram of Oriented Gradients (HOG) features from videos. The proposed model uses an SVM to classify 10 different expressions.

As a recent work, [84] presents a multimodal framework considering global hand locations/motions and local hand gesture details. By proposing a novel continuous sign language recognition framework that consists of a Hierarchical Attention Network with Latent Space (LS-HAN), the accuracy of continuous sign language recognition on the benchmark dataset of RWTH-PHOENIX-Weather 2014 was increased. [81] presents the most recent multimodal vision-based continuous sign language recognition study. The authors utilize a spatial-temporal multimodal network and design a joint optimization strategy to achieve end-to-end sequence learning. By training the proposed model on three benchmark datasets (RWTH-PHOENIX-Weather 2014, CSL, and RWTH-PHOENIX-Weather T), the computational results achieve new state-of-the-art accuracy and performance.

In addition to the vision-based model, some research utilizes special sensors to address multimodal sign language recognition such as [7] and [85]. [7] presents one of the most recent articles combining facial expression and hand gesture recognition. Both facial expressions (captured by Kinect) and hand gestures (captured by Leap motion) are considered simultaneously to increase the information that results in enhancing recognition performance. The HMM method is used for the sake of the recognition task, while an independent Bayesian classification combination approach is applied to improve the recognition performance. In order to illustrate the results, the authors collected a dataset containing 51 words in Indian sign language (which is publicly available now) and showed that the accuracy rate for the single hand gesture recognition was 96.05% for single and 94.27% for double hand gestures. [85] presents a novel multimodal framework considering both signals using two different sensors for face and hand gestures. A survey on both manual and non-manual sign language recognition systems is presented by [13].

2.1.4 Sign Language Translation

In this section, research that address translation of sign language to other form of languages is considered. These articles may have considered the sign language recognition problem to recognize sign language as a step before translating the sign language.

2.1.4.1 Within a language (Translating a sign language to text or speech)

As the first application of sign language recognition on a mobile platform, [86] proposed a sign language recognizing and translating platform to translate American sign language to text or speech. In their article, the proposed framework is able to recognize 16 static alphabets of ASL with an accuracy rate of 97.13% on average. Canny edge detection and region growing techniques are used to detect hand gestures from moderately complex backgrounds. In addition, SURF and SVM methods are used for the feature extraction and classifying. It is also demonstrated that the computational results are highly dependant of the illumination and background conditions, and the performance can decrease if any such kind of changes occur.

Not using machine learning or deep learning techniques, [87] developed a software-based sign language converter in which a geometric matching algorithm is used to recognize the cue symbols of ASL. Next, the values given by the matching algorithm after the decision making process are given to a case structure which generates a distinct text output for each corresponding match.

Finally, the text or its converted audio is generated for the final user.

[88] introduces a sign language converter system that recognizes static hand gestures of Indian sign language and translates them to texts in Indian spoken language. The proposed system captures images of the signs and uses Hue, Saturation, and Value (HSV) of the RGB colored images for hand tracking and segmentation (feature extraction). The captured images are then sent to a web hosting server and are given as input to the neural network in Matlab. After mapping the image to its equivalent in the Indian spoken language in Matlab, the converted text is sent back to the user's mobile device. To better illustrate the aforementioned process, an Android application using Android Studio is developed. One weak point of its research is its sensitivity to the lightening variations. The proposed system images are captured under constrained environments such as dark backgrounds, while in the real-world, the lightening conditions and background situations may vary.

Another study by [89] presents a real time gesture recognition system for recognizing sign language and converting it to speech output. The hand gesture recognition stage follows three steps: image prepossessing, feature extraction, and gesture classification. The former step uses Local Binary pattern, while the second step uses Gray level Co-occurrence Matrix. The last step of classification task is done by using a KNN classifier. Using a dataset containing 10 alphabets of ASL, the proposed method achieved an accuracy rate of 92.5% for k=3 and 94% for k=5.

Another way to approach the translation problem is converting a speech or text to signs for hearing impaired people. One of the recent papers that addresses this problem is [90], in which the authors use an automatic speech recognition module (ASR) to convert speech to text and then integrate it with a model which translates the text scripts to Arabic sign language (ArSL). To achieve this goal (Arabic text-to-sign (ArTTS) translation), 3D Avatar signers are generated and used. The 3D avatar character is able to employ four parameters in hands including hand shape, location, orientation, and movement. The results on testing 10 words demonstrate that the model is successful in translating the meaning of text scripts and has a detection rate as high as 85%.

2.1.4.2 Between languages (Translating a sign language to another)

This subsection is the literature review of sign language translation between languages, which in other words means translating a sign language to another sign language or another language's speech or text. Many deaf people do not know spoken languages and are not able to write or read spoken language [52], therefore, translating a sign language to a spoken language may not help the deaf community. As a result, another form of sign language translation may be needed, in which one sign language is translated to another one. At the same time, this area may consider translation of one sign language to its spoken language and then to another spoken one.

Table 2.1: An overview of some hardware-based sign language data sets

Authors	Name	Language	Gesture Type	Language level	Classes	Data Type
[64]	UCI Australian Auslan Sign Language dataset[1]	Australian	Dynamic	Alphabets	95	Data Glove
[17]	ASL Finger Spelling A [2]	American	Static	Alphabets	24	Depth Images
[17]	ASL Finger Spelling B [3]	American	Static	Alphabets	24	Depth Image

[1] Stands for Australian Sign Language (Available at: `https://archive.ics.uci.edu/ml/datasets/Australian+Sign+Language+signs+(High+Quality)`)

[2] Available at: `http://empslocal.ex.ac.uk/people/staff/np331/index.php?section=FingerSpellingDataset`

[3] Available at: `http://empslocal.ex.ac.uk/people/staff/np331/index.php?section=FingerSpellingDataset`

Authors	Name	Language	Gesture Type	Language level	Classes	Data Type
[91]	MSRGesture 3D [4]	American	-	Words	12	Depth Video
[25]	CLAP14	Italian	-	Words	20	Depth Video
[92]	ChaLearn LAP IsoGD[5] & ConGD[6]	-	Static & Dynamic	-	249	RGB & Depth Video
[93]	PSL Kinect 30[7]	Polish	Dynamic	Words	30	Kinect Video
[19]	ISL[8]	Indian	static	Alphabets, Numbers, Words	140	Depth Images

2.1.5 Sign Language Datasets

As previously discussed in subsection 2.1.1, conducted research in sign language recognition systems can be categorized in two main groups: hardware-based and vision-based recognition systems. Hardware-based recognition systems use datasets that are collected utilizing special colored gloves [14, 15, 94], special sensors, and/or depth cameras (such as Microsoft Kinect and Leap Motion) [16, 17, 18, 19, 95, 96, 97] to capture special features of the signer's gestures. Some well known hardware-based sign language datasets are listed in Table 2.1.

[4] Available at: `https://www.uow.edu.au/~wanqing/Datasets`

[5] Available at: `http://www.cbsr.ia.ac.cn/users/jwan/database/isogd.html`

[6] Available at: `http://www.cbsr.ia.ac.cn/users/jwan/database/congd.html`

[7] Available at: `http://vision.kia.prz.edu.pl/dynamickinect.php`

[8] Available at: `https://github.com/zafar142007/Gesture-Recognition-for-Indian-Sign-Language-using-Kinect`

Although utilizing hardware eases the process of capturing special features, they limit applicability where such hardware is not available. Therefore, vision-based sign language recognition systems utilizing datasets collected by regular cameras are proposed [3, 20, 98, 21, 22]. Table 2.2 lists some of the well known vision based sign language datasets.

Table 2.2: An overview of some vision-based sign language data sets

Authors	Name	Language	Gesture Type	Language level	Classes	Data Type
[60]	-	American	Static	Alphabets, numbers	36	Image
[17]	ASL Finger Spelling A [9]	American	Static	Alphabets	24	Image
[99]	HUST-ASL[10]	American	Static	Alphabets, Numbers	34	RGB & Kinect Image
[22, 100]	Purdue RVL-SLLL ASL Database[11]	American	-	Alphabets, Numbers, Words, Paragraphs	104	Image, Video
[101]	Boston ASLLVD[12]	American	Dynamic	Words	>3300	Video

[9] Available at: http://empslocal.ex.ac.uk/people/staff/np331/index.php?section=FingerSpellingDataset

[10] Stands for Huazhong University of Science & Technology

[11] Available at: http://www2.ece.ohio-state.edu/~aleix/ASLdatabase.htm and https://engineering.purdue.edu/RVL/Database/ASL/asl-database-front.htm

[12] Stands for American Sign Language Lexicon Video Dataset (Available at: http://www.bu.edu/av/asllrp/dai-asllvd.html)

Authors	Name	Language	Gesture Type	Language level	Classes	Data Type
[23]	ASL-LEX[13]	American	-	Words	Nearly 1000	Video
[3]	MS-ASL [14]	American	Dynamic	-	1000	Video
[102]	-	Arabic	-	Words	23	Video
[21]	RWTH-PHOENIX-Weather 2012[15]	German	-	Sentence	1200	Image
[20, 103]	RWTH-PHOENIX-Weather Multisigner 2014[16]	German	Dynamic	Sentence	¿1000	Video
[104]	SIGNUM[17]	German	-	words, Sentences	450 Words, 780 Sen-tence	Video
[15]	LSA16[18]	Argentinian	-	Alphabets, Words	16	Image
[15]	LSA64[19]	Argentinian	-	Words	64	Video
[19]	the ISL dataset[20]	Indian	static	Alphabets, Numbers, Words	140	Image

[13] Available at: `http://asl-lex.org/`

[14] Available at: `https://www.microsoft.com/en-us/download/details.aspx?id=100121`

[15] Available at: `https://www-i6.informatik.rwth-aachen.de/~forster/database-rwth-phoenix.php`

[16] Available at: `https://www-i6.informatik.rwth-aachen.de/~koller/RWTH-PHOENIX/`

[17] Available at: `http://www.phonetik.uni-muenchen.de/Bas/SIGNUM/`

[18] Available at:`http://facundoq.github.io/unlp/lsa16/index.html`

[19] Available at: `http://facundoq.github.io/unlp/lsa64/index.html`

[20] Available at: `https://github.com/zafar142007/Gesture-Recognition-for-Indian-Sign-Language-using-Kinect`

Authors	Name	Language	Gesture Type	Language level	Classes	Data Type
[105]	ISL hand shape dataset[21]	Irish	Static & Dynamic	-	23 Static & 3 Dynamic	Image Video
[106]	Japaneese Finger spelling sign language dataset	Japan	-	-	41	Image

2.2 Student Retention Literature

Much research has been conducted to identify and study factors affecting the post-secondary enrollment, persistence, completion, and graduation rates for deaf students [39, 107, 108, 109, 110, 111, 112, 112]. These studies show that cognitive factors such as academic preparation and English literacy are the important in post-secondary enrollment and success rate predictions of deaf students [43], but not for completion or graduation. These findings indicate that deaf students with adequate academic skills are still likely to drop out of college [113]. Hence, after deaf students enrol; in a post-secondary institution, other factors such as personal and non-cognitive factors are considered as stronger predictors for academic persistence and graduation rates [114, 115, 116, 117]. Some important non-cognitive factors influencing post-secondary persistent of deaf students are considered as academic experience, social integration, social skills, and clear goals and strategies.

Positive academic experiences such as having informal mentorship from faculty, participating in college activities outside of class, and collaborating with academic advisor, as well as high levels of social integration, such as being satisfied with social experience and having the ability to adjust socially, have a direct influence on students who persist post-secondary education [42, 118].

[21] Available at: `https://github.com/marlondcu/ISL`

Another important non-cognitive factor is social skills which include a high level of involvement in social activities, and the ability to perceive social situations and respond to the behaviors of others [108, 119]. Last but not least, having clear goals and strategies helps deaf students to have self-confidence and the desire to overcome post-secondary barriers [120].

Agent-based modeling (ABM) is a computational method that allows us to create, analyze, and model a system composed of autonomous decision-making artificial entities called agents [121]. ABMs are usually used in cases of modeling real-world phenomena that need more generalized models which can adapt to our world. ABMs can be coupled with other well developing methods such as machine learning –an area of artificial intelligence that attracted attentions in various fields of research such as cyber security [122] and computer vision [24]– to alter and enhance the way we analyze all different kinds of data.

An agent in ABM is an artificial autonomous individual who has properties, actions, and goals, which enables it to assess situations and make decisions based on defined rules [123]. In ABMs, agents may have interactions with themselves, other agents, and/or environments that permits us to execute and study how rules of agent behavior give rise to the emergence of macro-phenomena as the simulation output [124]. This capability of ABMs in capturing emergent macro-phenomena, along with other benefits in providing a natural description of a system and flexibility, has made them a popular modeling approach in various fields of research [125].

Deaf student post-secondary persistence is affected by various factors which if modified, can lead to a better life by providing the necessary boost for education and employment goals [39, 107, 112]. Such important phenomena can be best understood by using a bottom-up approach; ABM. Despite previously conducted research which identified non-cognitive factors influencing the post-secondary persistence of deaf students, interactions between these factors and their influence in predicting the persistence and graduation rate is still a barely explored field of research. In this work, we present an agent-based computational model for post-secondary persistence of deaf students with the goal of studying the effects of non-cognitive factors in post-secondary persistence. Based on the literature, academic experience, social integration, social skill, and clear goals and strategies described before are considered as four non-cognitive factors influencing post-secondary persistence for deaf students. To the best of our knowledge, this is the first agent-based modeling simulation for measuring the influence of non-cognitive factor on post-secondary persistence in deaf students.

CHAPTER 3: METHODOLOGY AND FRAMEWORK OF FACIAL EXPRESSION RECOGNITION AND HEARING IMPAIRED STUDENTS' RETENTION MODEL[1]

Vision-based recognition systems usually require large amounts of annotated data and vision-based sign language recognition systems are not an exception. Without more robust data concerning multi-modal features within sign language recognition, research and practical use for this concept remains stunted. Therefore, as the first step in proposing a multi-modal vision-based sign language recognition framework, the model requires a proper multi-modal sign language dataset.

To the best of our knowledge, although there are sufficient hand shape datasets in the context of sign language, the literature lacks facial expression datasets in the context of sign language or a multi-modal sign language datasets considering both hand shapes and facial expressions. The scarcity of multi-modal sign language datasets limits researchers' ability to study and propose multi-modal sign language recognition models that consider both facial expressions and hand gestures.

Therefore, in order to contribute to this research gap, we present a facial expression dataset (Facial expression Phoenix (FePh)) for a well-known continuous sign language dataset with full frames and hand shape images and annotations. Methodology of the aforementioned step is further discussed as following.

[1]Some parts of this chapter's material have been previously published in the International Journal of Electronics and Communication Engineering (2021), 15(3), 131 - 138 and 2020 Conference of the Computational Social Science Society of the Americas (CSS 2020).

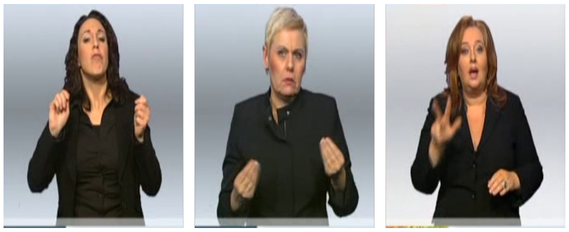
3.1 Facial expression Phoenix (FePh) Data collection

In this section we illustrate the Facial expression Phoenix (FePh) data collection's methodology, which to the best of our knowledge is the first annotated vision-based publicly available sequenced facial expression dataset in the context of sign language. The introduction of FePh in conjunction with RWTH-PHOENIX-Weather 2014 and RWTH-PHOENIX-Weather 2014 MS Handshapes datasets constitute the first sign language data with both handshape and facial expression labels. This characteristic enables us to propose novel multi-modal vision-based sign language recognition frameworks which consider two modalities of facial expressions and hand shapes.

In order to annotate facial expressions of a sign language dataset with annotated hand shapes, we considered the well-known publicly available continuous RWTH-PHOENIX-Weather 2014 dataset. The annotated hand shape dataset of the RWTH-PHOENIX-Weather 2014 is publicly available as RWTH-PHOENIX-Weather 2014 MS Handshapes dataset [66]. Therefore, by providing facial expression annotations for the same dataset, we introduce the first multi-modal sign language dataset. This enables us to utilize a multi-modal sign language dataset with both hand shape and facial expression annotations.

To achieve the aforementioned goal, we collected the full frame images of RWTH-PHOENIX-Weather 2014 development set that are identical to the RWTH-PHOENIX-Weather 2014 MS Handshapes dataset [66] and automatically detected, tracked, and cropped faces of all full frame images using facial recognition techniques. The result was a collection of cropped facial expression images. Figure 4.4 illustrates exemplary full frame and cropped facial images of the RWTH-PHOENIX-Weather 2014 and FePh datasets.

Twelve annotators (six women and six men) between 20 to 40 years old were asked to annotate the data. We asked annotators to answer three questions about each static image: the signer's



(a) Exemplary full frame images of the RWTH-PHOENIX-Weather 2014 dataset

(b) Exemplary cropped facial images of the FePh dataset using the full frame images of the RWTH-PHOENIX-Weather 2014 dataset

Figure 3.1: Exemplary full frame and cropped facial images of the RWTH-PHOENIX-Weather 2014 dataset. The cropped facial images shown in 3.1b are considered as Facial expression Phoenix (FePh) dataset images.

emotion, visibility, and gender. In terms of emotion, annotators could choose one or more of the following applicable basic universal facial expressions for each static image: ”sad”, ”surprise”, ”fear”, ”angry”, ”neutral”, ”disgust”, and ”happy”. Although more than seven emotions and their primary, secondary, and tertiary dyads exist, considering all of them was not within the scale of this project. Therefore, we offered the eighth class of ”None” as well. Annotators were asked to choose the ”None” class when none of the aforementioned emotions could describe the facial expression of the image. In addition, since annotators could choose more than one facial expression for each individual image, the combinations of basic universal facial expressions were also considered (interestingly, this did not result in choosing more than two emotions for each image) and shown by a ”_” in between such as surprise_fear. The sequence of emotions is not important in the secondary and tertiary dyads (i.e., surprise_fear and fear_surprise are the same).

With regard to the second question, visibility, we asked the annotators to evaluate whether the signer’s face is completely visible. Although the signer’s face was visible in majority of images, this was not always the case. The partial visibility of the face was due to the signer’s head movement, position, hand movement, and transitions from one emotion to another emotion. This helped us to detect and opt out these obscured images in the data. Figure 3.2 shows some obscured exem-

Figure 3.2: Exemplary images of obscured faces

plary images.

The last question of signer's gender was asked to provide statistics of signers' gender. This statistics enables future research in the effects of gender in expressing and detecting emotions and facial expressions.

For our labelling purpose, we took advantage of the Labelbox [126] annotating solution tool through which we defined an annotation project and randomly distributed images to be labeled by the annotators. In addition, due to the complexities of the facial images of the RWTH-PHOENIX-Weather 2014 dataset, we used the auto consensus option of the Labelbox tool. These complexities are as follows:

- The ambiguity of images, due to signer's movement, head position, and transitions from one emotion to another (e.g., eyes are closed and/or the lips are still open).
- Low quality (resolution) and blurriness of images.
- Mouthed words that confuse facial expression annotators.
- Personal differences between signers expressing facial expressions.
- The best facial expression that describes the image is not included in the dataset.
- Images may not be in facial expression's top frame.

- Large intra-class variance (such as "surprised" emotion with open or closed mouth).
- Inter-class similarities.

With the usage of the auto consensus option of Labelbox, we asked more than one annotator (i.e., three annotators) to annotate about 60% percent of the data. For the images with three labels, we chose the most voted emotion as the final label of the facial image. In cases where there was not a most voted emotion, but the image was a part of a sequence of images, we have assigned labels based on the before or after images' facial expression of the same sequence. On the other hand, if there was not a most voted emotion, and the image was not a part of a sequence of images (i.e., one single image without any sequence), we asked our annotators to relabel the image. In this case, all images needed to be labelled by three different annotators.

3.1.1 Facial expression Phoenix (FePh) Data Usage

The FePh facial expression dataset produced with the above method is stored on Harvard Dataverse (`https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/358QMQ`). All facial images are stored in "FePh_images.zip". Although the full frame images of the FePh dataset are identical to the RWTH-PHOENIX-Weather 2014 images, the images' filename in the FePh_images.zip are different from the original images. FePh filenames consider both image folder_name and image_file_number. For example, the full frame image of the facial image with filename "*01August_2011_Monday_heute_defa ult-6.avi_pid0_fn000054-0.png*" in the FePh dataset is identical to the full frame image with the directory of "*... / 01August_2011 _Monday_heute_default-6 / 1 / 01August_2011 _Monday_heute.avi_pid0_ fn000054-0.png*" in the RWTH-PHOENIX-Weather 2014 and "*... / 01August_2011 _Monday_heute_default-6 / 1 /_.png_ fn000054-0.png*" image in the RWTH-PHOENIX-Weather 2014 MS Handshapes dataset. This helped us to store all images in one single folder (FePh_images). The FePh_labels.csv file contains

Table 3.1: The facial expression labels with their corresponding numbers

label#	Emotion	label#	Emotion	label#	Emotion
0	neutral	10	anger_neutral	52	sad_disgust
1	anger	21	disgust_anger	53	sad_fear
2	disgust	31	anger_fear	60	neutral_surprise
3	fear	32	fear_disgust	61	surprise_anger
4	happy	40	neutral_happy	62	surprise_disgust
5	sad	50	neutral_sad	63	surprise_fear
6	surprise	51	sad_anger	65	surprise_sad
7	none				

images' filenames, facial expression labels, and gender labels. To ease data usability, we stored the facial expression labels as codes. Table 3.1 shows the facial expression labels with their corresponding code numbers. In addition, 0 and 1 in the gender column represent the male and female genders, respectively.

3.2 Hearing Impaired Students' Retention Agent-based Model

Among several student retention models, the one presented by Tinto [1, 127] is held in high regard and is the most cited model [40, 128]. Tinto's model, shown in Figure 3.3, provides a heuristic and theoretical framework for understanding student behaviour while describing the factors influencing the persistence process. With some modifications, this model can be applied to deaf college students as well [39, 107]. According to this theory, a combination of student characteristics and academical, environmental, and social integration in an institution influence the student's departure decision. To create our model, we use the theoretical model of Tinto [127]. By only considering the influence of four non-cognitive factors (academic experience, social integration, social skill, and clear goals) on Tinto's model, we created a new simplified framework as shown in Figure3.4. This figure shows the diagram of four non-cognitive factors influencing student departure decision

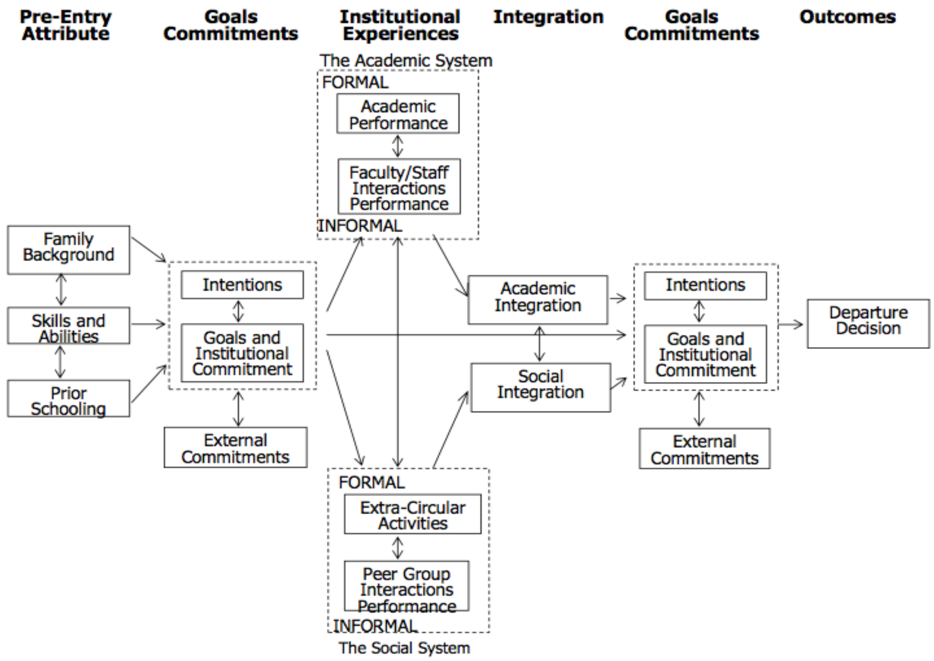


Figure 3.3: A conceptual diagram for dropout from college presented by [1]

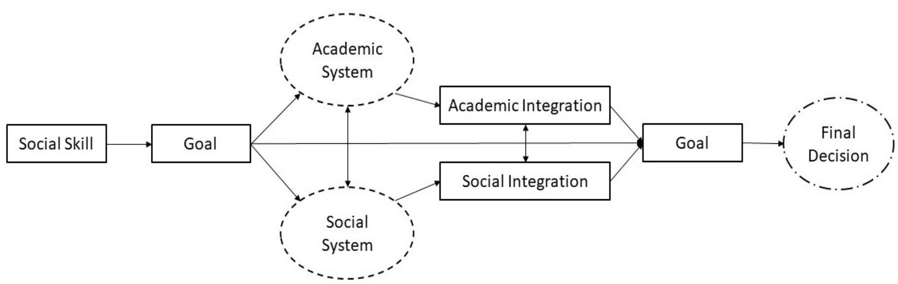


Figure 3.4: Diagram of four non-cognitive factors effecting student's departure decision based on Tinto's dropout from college model

utilizing the same relations as Tinto's model (Figure3.3). In this work, this theoretical framework is used to create the ABM and measure the probability of the departure decision of a student.

Our proposed ABM model involves two categories of actors: teachers and deaf agents. Deaf agents are members of the general population who have graduated from secondary school and are capable of attending post-secondary education regardless of their age. While teachers stay at college during college years (runs of the model), deaf agents can decide whether to attend college or not. The specifications, attributes, and behavioral rules of teachers and deaf agents during the SETUP and RUN phases of the ABM are as follows;

SETUP Specifications. In order to setup the model, the user sets exogenous parameters of the total number of agents available in the model (called "*num_agents*"), fraction of teachers (called "*frac_teachers*"), and college attendance percentage of deaf students (called "*College_Attendance*"). The total number of agents will show the sum of teachers and deaf agents. To show the mathematical equations, the total number of teachers and deaf agents is equal to:

$$num\_teachers = num\_agents * frac\_teachers$$

$$num\_deaf\_agents = num\_agents - num\_teachers$$

In the model settings, academic experience, social integration, social skill, and clear goals are named as *Academic_Experience*, *Social_Integration*, *Social_Skill*, and *Goal*, respectively. Clear goals and social skill factors are considered to be exogenous and heterogeneous across agents. Lacking real-world data, each agent's value for the two exogenous factors are considered to have the uniform distribution on the interval of (0,1); $U(0,1)$. Setting the value of 0 for each of these factors show the lowest level of the factor. As the value increases, the level of having that positive factor increases, while "1" is the highest level of the factor.

On the other hand, the other two factors, academic experience and social integration, (which are usually dependent to connections and interactions with others, such as teachers and other students

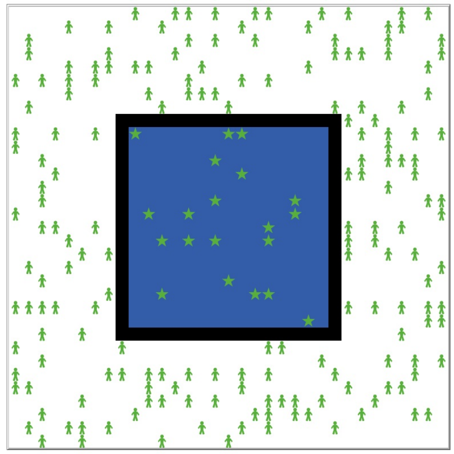


Figure 3.5: Exemplary visualizations of the setup screen of the student retention ABM. The blue square in the middle of the screen with a thick black edge shows the college. White locations surrounding the college are considered residential locations. The green human shapes on the residential locations and the green stars at college represent deaf agents and the teachers, respectively.

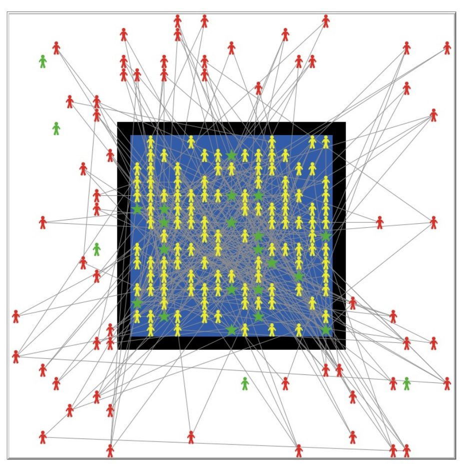


Figure 3.6: Exemplary visualizations of the proposed NetLogo model after two ticks of running the model. Screen of the student retention ABM while running the model. As soon as deaf agents attend college at the first tick (year), their color changes to yellow and they will be called students. If students quit the college, their color turns to red and they will leave college and move into a random location on the residential locations. The gray lines connecting agents show the created links of each student with teachers and other students.

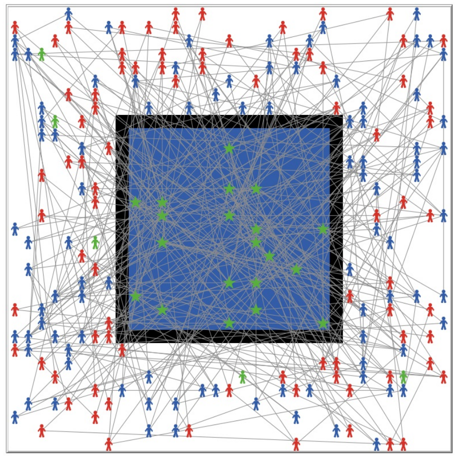


Figure 3.7: Exemplary visualizations of the proposed NetLogo model after four ticks of running the model. Screen of the student retention ABM after four ticks. At the end of the fourth year, students will graduate. By graduating, their color will turn to blue and they will move into one of the residential locations (leave college).

Table 3.2: Input assumptions for all runs and experiments. Factors values changes with steps of 0.1 in the interval of (0.1,1).

factor influence	num_agents	college_attendance	repetitions	goal	academic experience	social skill	social integration
goal	200	87.2	10	(0.1,1)	0.5	0.5	0.5
academic experience	200	87.2	10	0.5	(0.1,1)	0.5	0.5
social skill	200	87.2	10	0.5	0.5	(0.1,1)	0.5
social integration	200	87.2	10	0.5	0.5	0.5	(0.1,1)

at the college), are considered to be derived from a multiplication of the initial value of the factor that the user sets and the number of the links (connections) with others that a student create at college.

There are two locations in our ABM screen: college, and residential. College is an institution that deaf agents can attend to pursue their post-secondary education. College is located at the middle of the screen and is colored in blue with a thick black line around outside edges. On the other hand, residential locations can be any other kind of location in which deaf people cannot have post-secondary education. Residential locations are represented by the white color surrounding the college. Figure 3.5 shows both college and residential locations. When the model starts, only teachers are located at the college; all deaf agents are randomly located in one of the empty spaces of the residential locations. Figure 3.5 shows the setup screen of the proposed ABM. In Figure 3.5, deaf agents are shown as green human shapes located at white residential locations while teachers are shown as green stars located randomly in the college.

RUN Specifications. To begin each run of the model, the user sets the initial number of agents, fraction of teachers, and the college attendance percentage as well as the initial values of the four non-cognitive factor variables. The model runs based on ticks (i.e,., each run ticks the model once). For simplification, in this work we only consider 4-year colleges or university programs (that we call ”college” at the rest of this dissertation) for post-secondary education. Therefore, model runs for the total of four ticks to achieve the end of the fourth year at college. At the end of the fourth year, students graduate and leave the college.

The deaf agents can only decide whether to attend college or not in the first tick (that is, the first year). If they do, they move to one of empty spots at the college location and their color changes to yellow. Otherwise, they will remain at residential locations without any color change. We call the attended college agents ”students”. Students are deaf agents who not only attended college, but also are still persisting at college at the end of each tick (i.e., year). If attended agents decide to depart college during a tick, they move out of the college and move to a random spot in the residential locations. These agents who attended college but departed it are called ”quitters”. Quitters are shown as red human shapes located at residential locations. For simplicity, we assume that if students depart college, they do not attend it again.

In the second, third, and fourth ticks of the model, students can only decide whether to depart college or not. Just as in the first run, if they decide to depart, they will leave college and locate to a random spot of the residential locations. Their color will also turn to red. The screen of the model after running two ticks is shown in Figure 3.6.

During college years, students can randomly create links or connections (shown with gray lines in Figure 3.6 and 3.7) with teachers or other students at the college. The total number of these links are used to adjust and update the level of social integration and academic experience factors of the student. The number of links that a student creates with teachers and other students will be used to evaluate the level of the academic experience and social integration factors of the student, respectively. The more links, the higher the level of the factor. The number of links that each student creates with other teachers and students can vary between zero to three and zero to eight, respectively. The minimum value of both social integration and academic experience of each student is considered to be 0.2. Each created link with other students add 0.1 to the minimum value of a student's social integration. Similarly, each created connection with a teacher at the college adds 0.1 to the minimum value of the student's academic experience factor.

Similar to the first three ticks, during the fourth tick of the model, students can still persist in college. If they do, they have finished four years of college and therefore, they have completed the program have graduated. These students are called ”graduates”. Graduates, whose color change to blue, depart college at the end of the fourth year and move to one of the empty spots at the residential locations. Figure 3.7 shows the model screen after the fourth tick.

CHAPTER 4: FACIAL EXPRESSION PHOENIX (FEPH) TECHNICAL VALIDATION AND FINDINGS[1]

The FePh dataset is created by manually labelling 3359 images of the RWTH-PHOENIX-Weather 2014 development set that are identical to the full frame images of the RWTH-PHOENIX-Weather 2014 MS Handshapes dataset. Seven universal basic emotions of ”sad”, ”surprise”, ”fear”, ”angry”, ”neutral”, ”disgust”, and ”happy” are considered as facial expression labels. In addition to these basic emotions, we asked annotators to choose all the emotions that may apply to an image. This resulted in secondary and tertiary dyads of seven basic emotions such as fear_sad, fear_anger, etc. Interestingly, this did not result in having combinations of three basic emotions. Figure 4.1 shows the corresponding graph of seven basic emotions, their primary, secondary, and tertiary dydes presented in FePh dataset. Seven basic emotions are shown by colored circles with the emotion labels written inside them. Other colored circles connecting each two basic emotions illustrate the secondary or tertiary dyads of the basic emotions that are connected to. Emotion ”Happy” has only one dyad, which is with ”Neutral” emotion, named as ”Neutral_Happy”.

Although the FePh dataset presents annotated facial expression for all hand shape classes of the RWTH-PHOENIX-Weather 2014, we analyzed the facial expression labels for the top 14 hand shapes (i.e., classes ”1”, ”index”, ”5”, ”f”, ”2”, ”ital”, ”b”, ”3”, ”b_thumb”, ”s”, ”pincet”, ”a”, ”h”, and ”ae”). This is due to the demonstrated distribution of the counts per hand shape classes in [**?**] that shows the top 14 hand shape classes represent 90% of the data. Seven universal basic facial expressions and their secondary or tertiary dyads occur with different frequencies in the data. Figure 4.2a shows the distribution counts per facial expression class in the data. As it shows, about 90% of the data is expressed with basic facial expressions. In addition, Figures 4.2b and

[1]This chapter’s material has been previously published in the International Journal of Electronics and Communication Engineering (2021), 15(3), 131 - 138.

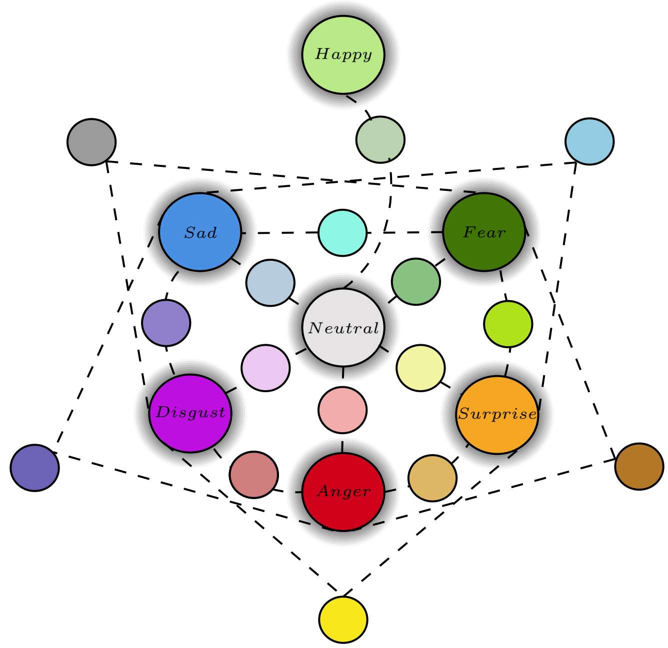
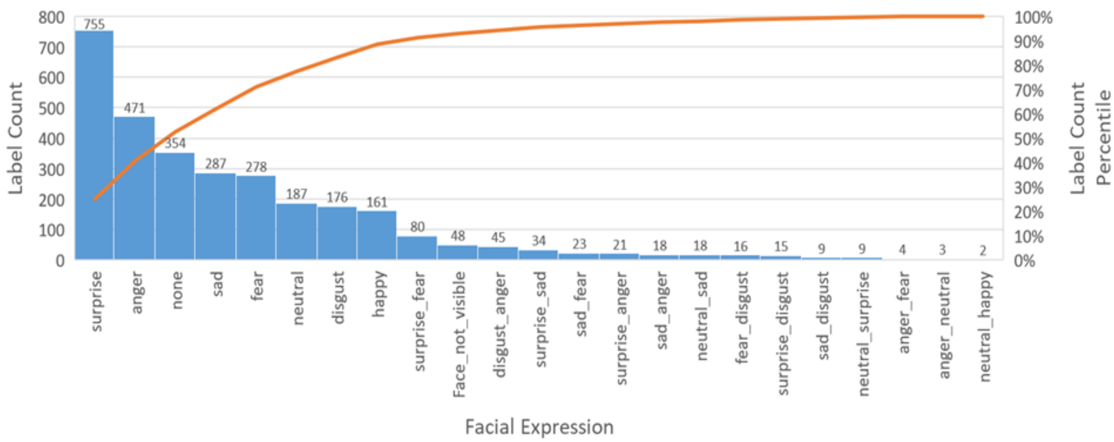


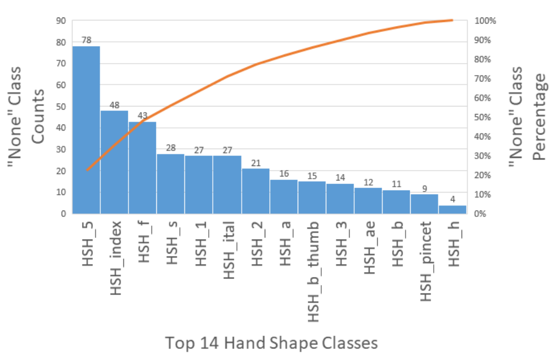
Figure 4.1: Graph displaying the primary, secondary, and tertiary dyads on seven basic universal emotions in the FePh dataset. Seven basic universal emotions are shown with colored circles containing their name. In addition, colored circles that are connecting each two basic emotions show the secondary or tertiary dyads of the basic emotions that they are connected to. For example, the orange circle connected to the "Disgust" and "Surprise" emotions, shows the emotion of "Disgust_Surprise". Emotion "Happy"'s only dyad is with "Neutral" emotion named as "Neutral_Happy".

4.2c illustrate the frequency of images with obscured faces and the "None" class in the top 14 hand shape classes, respectively.

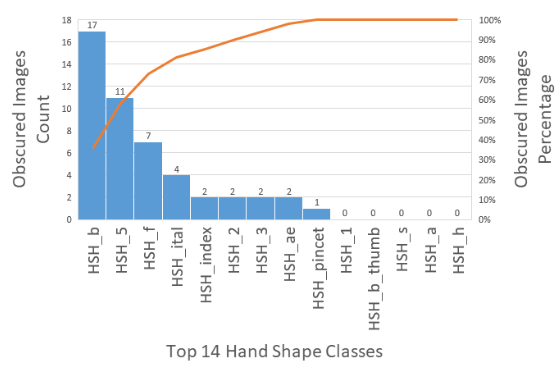
By analyzing facial expressions per hand shape class, we found out that more than one facial expression class represents each hand shape class. Figure 4.3 shows the frequency heatmaps of the seven facial expressions and their primary, secondary, and tertiary dyads for the top 14 hand shape classes. Each heatmap illustrates the frequency of facial expressions based on the facial expression graph of seven basic universal emotions and their primary, secondary, and tertiary dyads (shown in Figure 4.1) for one of the top 14 hand shape classes. The heatmaps show that more than one



(a) Pareto chart showing the distribution counts per facial expression class in the FePh dataset for the top 14 hand shape classes (briefly shown as HSH). As the chart shows, basic facial expressions represent 90% of the data.



(b) Pareto chart showing the distribution of the ”None” facial expression class counts per top 14 hand shape classes (HSH)

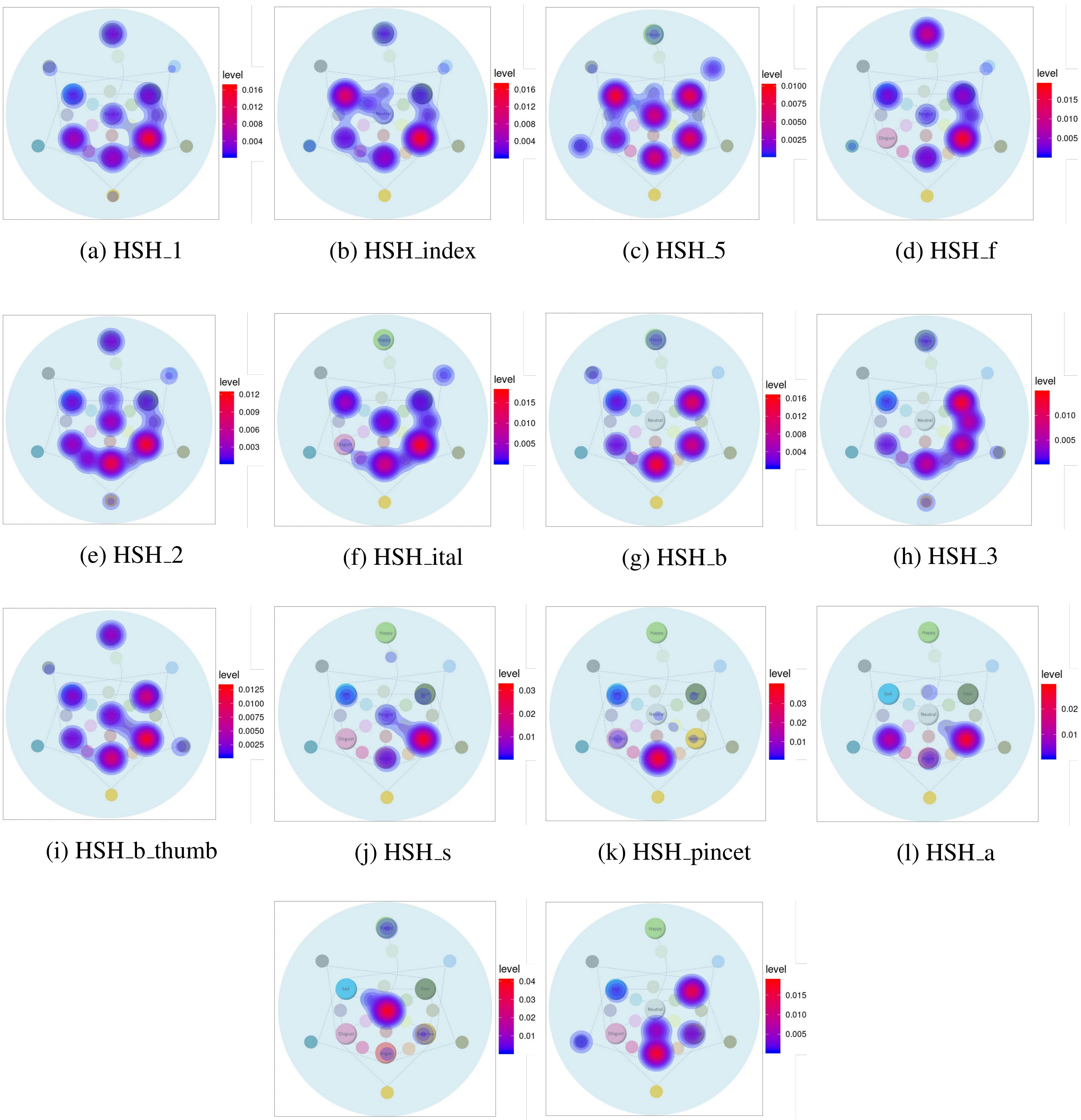


(c) Pareto chart showing the distribution of the obscured image counts per top 14 hand shape classes (HSH)

Figure 4.2: Pareto chart showing the distribution of the facial expression ”None” and obscured image counts for top 14 hand shape classes (HSH).

facial expression is expressed within a single hand shape class, which is due to the complexity of sign language in using facial expressions with hand shapes. Two of these complexities that affect performing different facial expressions within each hand shape class are as follows:

First, although some meanings are communicated using only one hand (usually the right hand),



(a) HSH_1 (b) HSH_index (c) HSH_5 (d) HSH_f

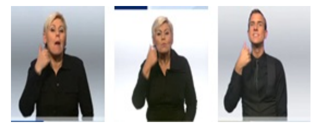
(e) HSH_2 (f) HSH_ital (g) HSH_b (h) HSH_3

(i) HSH_b_thumb (j) HSH_s (k) HSH_pincet (l) HSH_a

(m) HSH_h (n) HSH_ae

Figure 4.3: Heatmaps showing the frequency distribution of the top 14 hand shape classes. Each heatmap, assigned to one hand shape class (briefly mentioned as HSH), shows the frequency of facial expressions on the assigned hand shape class over the facial expression graph of seven basic universal emotions and their primary, secondary, and tertiary dyads. As they show, more than one facial expressions are expressed within a single hand shape class.

(a) Exemplary images of hand shape class ”1”



(b) Full frame images with similar hand shapes but different facial expressions

Figure 4.4: Exemplary full frame images of the RWTH-PHOENIX-Weather 2014 dataset

many sign language meanings are communicated using both hands with different hand poses, orientations, and movements. Figure 4.4a shows some exemplary full frame images of hand shape class ”1” of the RWTH-PHOENIX-Weather 2014 dataset with different facial expressions. As the figure illustrates, the usage of right hand shapes have large intra-class variance (i.e., the left hand may not be used or may perform similar or different hand shape from the right hand) that may affect the meanings, and as a result, the facial expressions corresponding to them. The first top row in the figure shows the full frame images with the right hand shape of class ”1” and no left hand shape. The second row shows some exemplary full frame images, in which the signer has used both hands. In this row, although both right and left hands demonstrate the same hand shape class (hand shape class ”1”), their pose, orientation, and movement can differ, which may affect the corresponding facial emotion. The third row of images in Figure 4.4a illustrates full frame examples of using both hands with different hand shapes and facial expressions. Therefore, although RWTH-PHOENIX-Weather 2014 MS Handshapes is a valuable resource presenting right hand shape labels, it lacks pose, orientation, and movement labels of the right hand along with the left hand shape labels. Adding this information to the data affects the communicated meanings as

Table 4.1: Correlation matrix of facial expressions' dummy variables and their frequencies for the top 14 hand shape classes. Empty cells show the absence of the correlation (the facial expression is not expressed in the hand shape class images).

	Frequency of facial expressions for hand shape class													
	1	index	5	f	2	ital	b	3	b_thumb	s	pincet	a	h	ae
Frequency	1	1	1	1	1	1	1	1	1	1	1	1	1	1
neutral	-0.027	-0.145	0.153	-0.210	0.115	0.066	0.666	-0.221	-0.019	-0.168	-0.150	-0.371	0.995	
anger	0.194	0.114	0.176	-0.047	0.582	0.405	-0.164	0.287	0.343	-0.093	0.991	0.087	-0.259	0.350
disgust	0.233	-0.024	-0.067	0.041	0.075	-0.231	0.459	-0.076	-0.099	-0.242	-0.112			
fear	0.064	-0.089	0.257	0.335	-0.126	-0.061	-0.277	0.697	0.303	-0.192	-0.150			0.250
happy	0.038	-0.080	-0.206	-0.036	-0.045	-0.241	-0.126	-0.124	0.022	0.850	-0.073		-0.209	
sad	-0.040	0.502	0.338	0.805	-0.032	0.130	0.308	-0.124	-0.019	0.379	-0.170			-0.350
surprise	0.857	0.737	0.234	0.139	0.662	0.787	-0.164	0.383	0.624		-0.015	0.612	-0.209	-0.250
none	-0.053	0.187	0.651		-0.005	0.034		0.093	0.263			0.546	-0.109	0.751
anger_neutral														-0.150
disgust_anger	-0.170	-0.153			-0.005	-0.199	-0.353	-0.196	-0.300					
anger_fear								-0.196	-0.260					
fear_disgust	-0.170		-0.230				-0.296		-0.300					
happy_neutral										-0.292				
neutral_sad		-0.097	-0.218										-0.209	
sad_anger		-0.169	-0.160	-0.298				-0.221	-0.300					-0.350
sad_disgust		-0.169	-0.195											
sad_fear		-0.145	-0.206		-0.139							-0.437		
neutral_surprise	-0.222								-0.260	-0.242	-0.150	-0.437		
surprise_anger	-0.209				-0.192	-0.178		-0.148						
surprise_disgust	-0.183				-0.219			-0.172						
surprise_fear	-0.118	-0.113	-0.241	-0.189	-0.192	-0.125		0.214						
surprise_sad	-0.190	-0.169		-0.287	-0.219	-0.178								

well as the facial expressions that are expressed.

Second, due to the communication of grammar via facial expressions, identical hand shapes may be performed with different facial expressions. Figure 4.4b demonstrates some images of this kind that despite the similarity of hand shapes, the facial expressions are different. This complex usage of hands with large intra-class variance and inter-class similarities help signers to communicate different meanings with similar or different facial expressions.

In addition to the above, the frequency of facial expressions expressed in each hand shape class shows evidence of a meaningful association between hand shapes and facial expressions in the data. To better illustrate this correlation, we calculated the correlation matrix of facial expressions' frequency in each hand shape class. Since the correlations between the facial expressions together is not the focus of this manuscript, the first column of each correlation matrix that shows the correlation between frequency and each facial expression is only considered. Table 4.1 illustrates the first columns of facial expressions and their frequencies of occurrence in the top 14 hand shape classes correlation matrices. Monitoring each column in Table 4.1 gives the most correlated facial expressions for each hand shape. For example, in the column of hand shape class ”3”, the positive values of 0.697, 0.383, 0.287, 0.214 and 0.093 (that are in intersections with ”fear”, ”surprise”, ”anger”, ”surprise_fear” and ”None”, respectively) show the positive correlation values with facial expressions in hand shape class ”3”. These highly correlated facial expressions in each hand shape class can also be interpreted from heatmaps illustrated in Figure 4.3.

4.1 Usage Notes

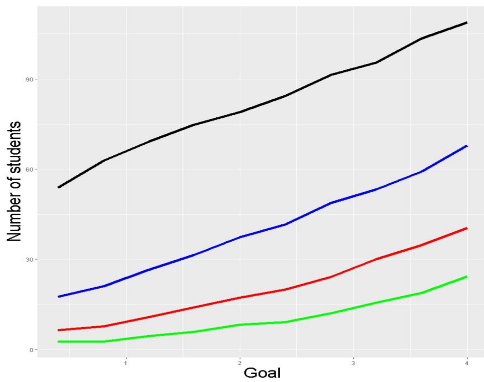
To the best of our knowledge, this dataset is the first annotated vision-based publicly available sequenced facial expression dataset in the context of sign language. Although the number of facial images is enough for statistical and some machine learning methods, it may not be sufficient for some of the state-of-the-art learning techniques in the field of computer vision. Therefore, for such studies, we suggest users to create matched samples choosing subjects from the dataset. This work not only provides an annotated facial expression dataset with different head poses, orientations, and movements, but also contributes in availability of a sign language dataset with both hand shape and facial expression labels with attributions in multi-modal future works in the field. In addition, this dataset has a wider application in other research areas such as gesture recognition and Human-Computer Interaction (HCI) systems.

CHAPTER 5: TECHNICAL VALIDATION AND FINDINGS OF POST-SECONDARY PERSISTENCE OF DEAF STUDENTS AGENT-BASED MODEL[1]

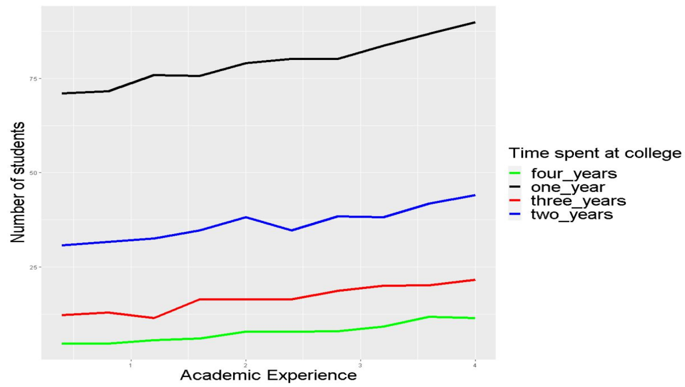
The implementation of this work is done using NetLogo 6.1.1 [129]. The results of replicating, repeating, and reproducing the results, with input assumptions for all runs are provided in Table 3.2.

We performed four experiments using the behavior search tool of NetLogo to find the influence of each factor on the college persistence of deaf students for years one, two, three, and four. Persistence of four years at college is considered as graduating from college and receiving a 4-year college degree. In order to run each experiment, we considered one non-cognitive factor value changing in an interval of (0,1) with steps of 0.1 while the values of the other three factors are fixed at 0.5. The results of these four experiments are shown in Figure 5.1. As the figure shows, although increasing the level of a factor increases the persistence of students at post-secondary education, these four non-cognitive factors' impacts vary from one to another. In addition, as all four figures show, the black line showing the number of students who persisted one year at college is further away from the other lines for students who persisted two or more years at college. This shows that the majority of students depart post-secondary education during the first year of education. In other words, students who do not have clear goals and strategies, positive academic experience, strong social skills, and high levels of social integration are most likely to depart post-secondary education during the first year. If they persist after the first year, the chance of departing decreases with a high margin.

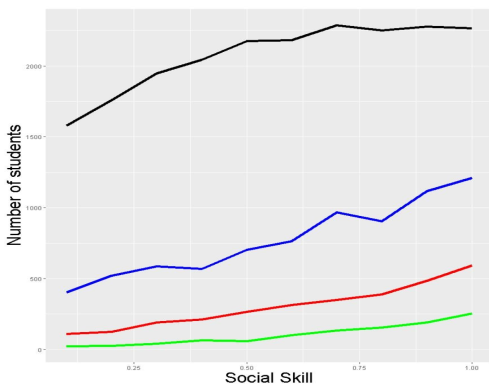
[1]This chapter's material has been previously presented in the 2020 Conference of the Computational Social Science Society of the Americas (CSS 2020).



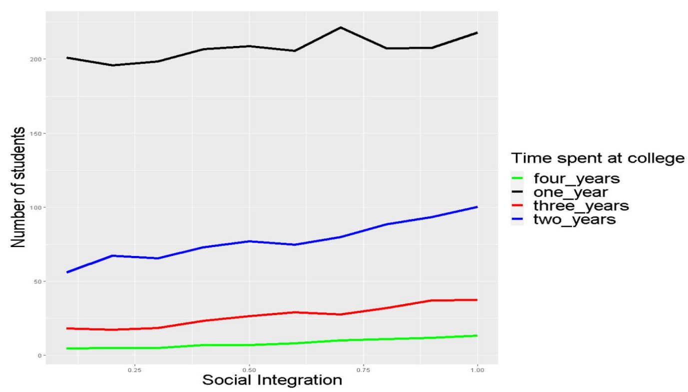
(a) Goal factor



(b) Academic experience factor



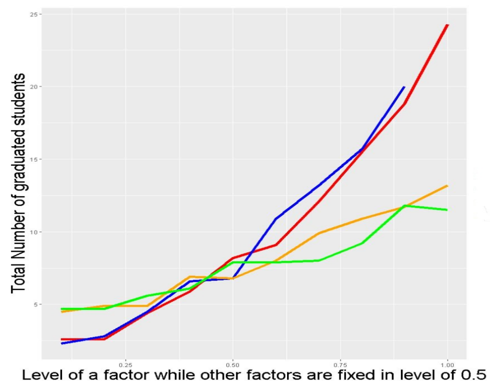
(c) Social skill factor



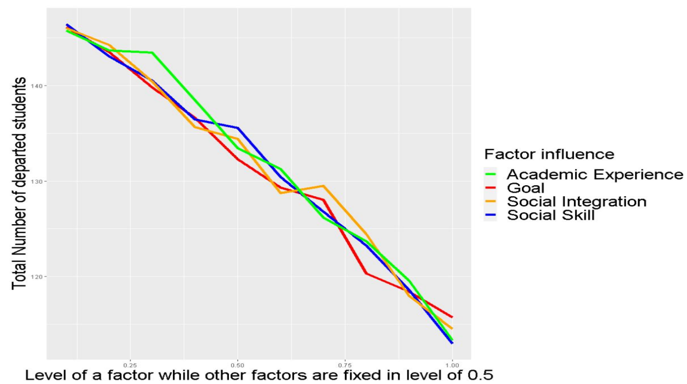
(d) Social integration factor

Figure 5.1: Results of behavior space experiment for four non-cognitive factors of goal, academic experience, social skill, and social integration. Each graph shows the result of behavior search for one factor changing in an interval of (0,1) with steps of 0.1. To run the experiment for one factor, all other three factors are considered to have a value of 0.5.

In order to compare the number of graduated and departed students with different levels of non-cognitive factors together in two separate graphs, an experiment in which the level of all factors can change in an interval of (0.1,1) with steps of 0.1 is performed. This enables us to extract a plot for the total number of graduated (i.e., all green lines of four_year at college from Figure 5.1 in one single plot) and departed students with different levels of factors. Figure 5.2 shows the graphs of the number of graduated and departed deaf students with fixed three non-cognitive factors at



(a) Number of graduated students based on different levels of one factor while fixing the other factors.

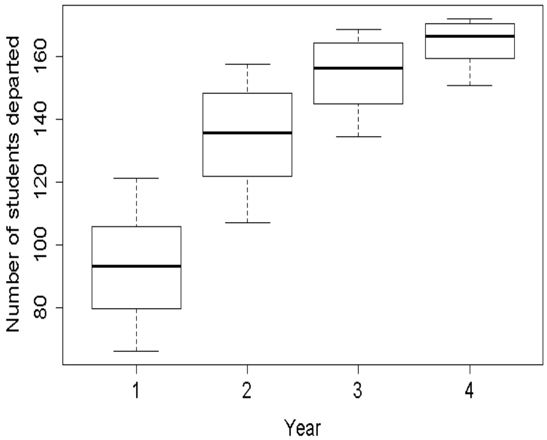


(b) Number of departed students based on different levels of one factor while fixing the other factors.

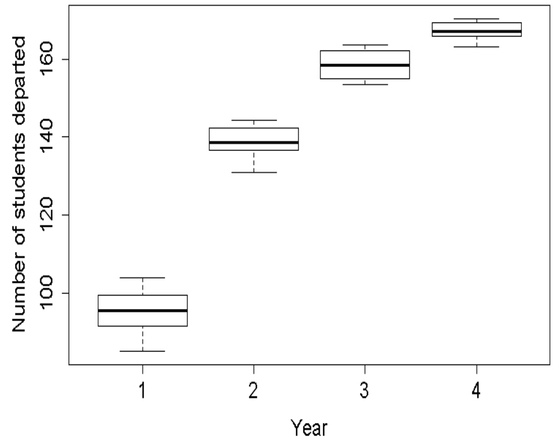
Figure 5.2: Number of graduated and departed students based on different levels of one factor while fixing the other factors to the level of 0.5 during four years of post-secondary education.

level 0.5 and varying one factor in an interval of (0.1,1) with steps of 0.1 during four years of post-secondary education. Different colored lines show the line plot for the assigned factor for them that is changing. As Figure 5.2b shows, although by increasing the level of factors the total number of departed students decreases, it still does not merge to zero. This indicates that despite the high levels of non-cognitive factors, some students may still depart college due to other factors, issues, or concerns. Furthermore, considering low levels of academic experience and social integration, the total number of departed students are noticeably more than other factors, while in contrast, with high levels of goal and social skill, the total number of departed students are noticeably decreased. This illustrates the importance of academic experience and social integration in low margins and strong goals and social skill in high margins when a student decides whether to depart college or not.

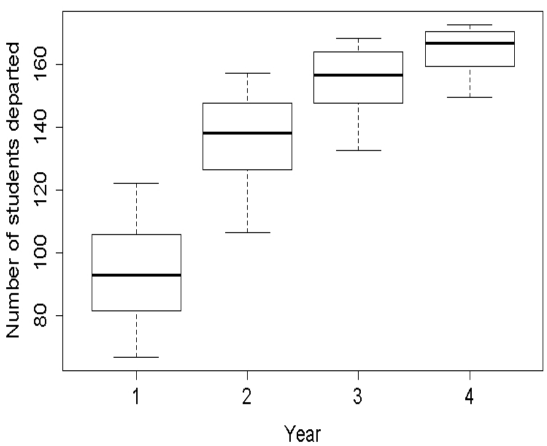
The results of the sensitivity analysis for each individual non-cognitive factor can be seen in Figure 5.3. The +- 10% range is illustrated using error bars while the boxplots represent the range of



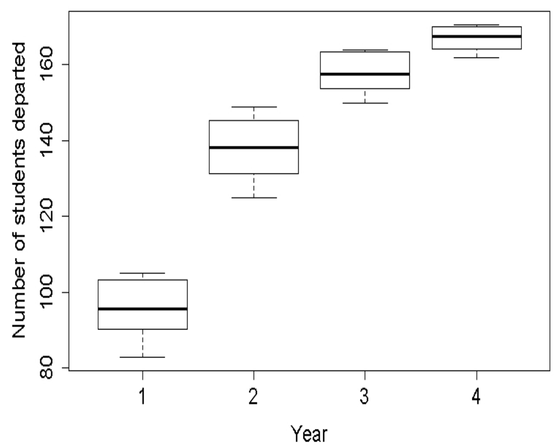
(a) Boxplot of the results of the behavior space experiment for goal factor



(b) Boxplot of the results of the behavior space experiment for academic experience factor

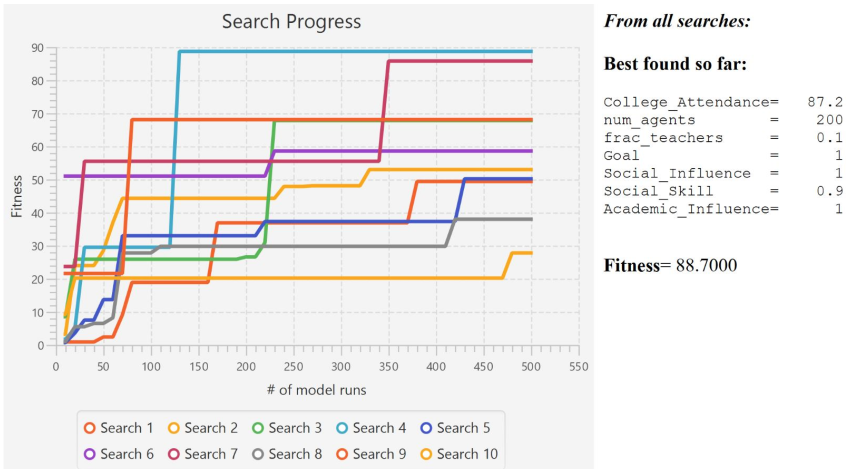


(c) Boxplot of the results of the behavior space experiment for social skill factor

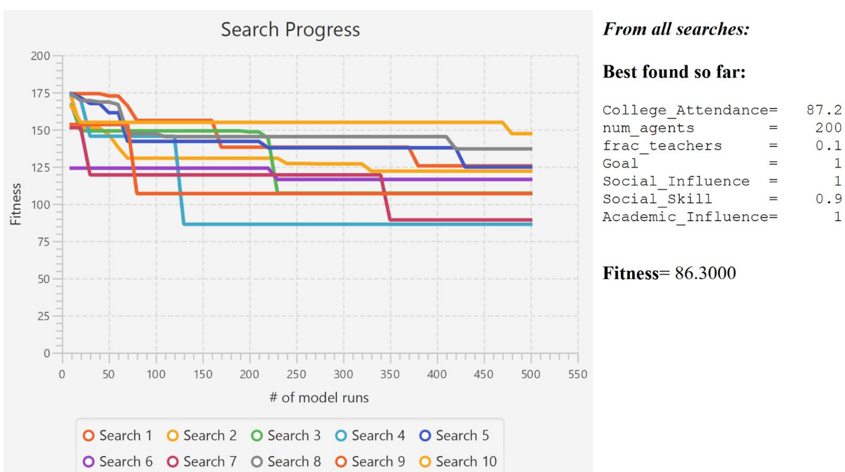


(d) Boxplot of the results of the behavior space experiment for social integration factor

Figure 5.3: Boxplots of the results of the behavior space experiment for the four non-cognitive factors of goal, academic experience, social skill, and social integration. Each graph shows the result of the behavior search for one factor changing from interval of (0,1) with steps of 0.1. To run the experiment for one factor, all other three factors are assigned the value of 0.5.



(a) Plot of ten different searches for maximizing the total number of departed students. Maximized fitness is achieved with level 1 for goal, social skill, and social integration and level of 0.8 for academic integration.



(b) Plot of ten different searches for minimizing the total number of departed students. Minimized fitness is achieved with level 1 for goal, social skill, and social integration, and level of 0.8 for academic integration.

Figure 5.4: Result screens of running the Behavior Search tool on the proposed NetLogo model

number of departed students in ten repetitions of the model. Although boxplot graphs of the four factors' influence are somewhat similar, it is apparent that the model is highly sensitive to the academic experience factor. In addition, as Figure 5.1 illustrates, the number of departed students between year one to four is not linear, and the slope decreases as the year increases.

To maximize the number of graduated students as well as to minimize the total number of departed students, we performed a behavior search experiment using the proposed NetLogo model. Figure 5.4 shows the results of the behavior search experiment for both aforementioned objectives. As the figure illustrates, the best value for both objectives is achieved by the value of 1 for goal, academic integration, and social integration factors, and value of 0.9 for social skill factor. The maximum number of graduated students as well as the minimum number of departed students based on the aforementioned factor values are 88.700 and 86.3, respectively. However, because we are counting the number of persons, both values will be rounded up to 89 and 87. These results show that almost half of deaf students decide to depart post-secondary education before graduation.

CHAPTER 6: CONCLUSION

In this dissertation, we first presented the FePh dataset, which to the best of our knowledge, is the first real-life annotated sequenced facial expression dataset in the context of sign language. FePh contains over 3000 sequenced images of the RWTH-PHOENIX-Weather 2014 dataset that are identical to the full frame images of the RWTH-PHOENIX-Weather 2014 MS Handshapes dataset. FePh in conjunction with RWTH-PHOENIX-Weather 2014 and RWTH-PHOENIX-Weather 2014 MS Handshapes datasets constitute the first sign language data with both handshapes and facial expression labels. We hope this unique characteristic will propel research in multi-modal sign language and gesture recognition.

Second, we studied the effects of four non-cognitive factors: having clear goals, social integration, social skills, and academic experience, on post-secondary persistence or retention of deaf students from an agent-based modeling (ABM) and simulation approach. To the best of our knowledge, we present the first ABM simulation for the aforementioned problem in order to simulate students retention behavior and discover the effects of non-cognitive factors in students persistence and departure decisions. Our results indicate that first year persistence at a 4-year post-secondary education (e.g., university, college) plays an integral role in student's persistence and graduation. In other words, if a student persists after the first year of a post-secondary education, the chances of student departure decreases with a high margin. In addition, the best persistent rate of the model is achieved by a social skill factor of 0.9 and other factors of 1. We believe that presenting and creating ABM brought significant benefits to studying deaf students' departure decisions during post-secondary education.

6.1 Future Work

As preliminary results and analysis of the FePh dataset indicate a meaningful relationship between two important modals of sign language (i.e., hand shapes and facial expressions), for future work, we propose applying multi-modal learning and computer vision techniques on joint RWTH-PHOENIX-Weather 2014 MS Handshapes and FePh datasets. We believe that the introduction of this dataset will allow the facial expression, sign language, and gesture recognition communities to improve their learning techniques to the latest levels of computer vision trends.

For the agent-based education model, we believe that presenting more sophisticated agent-based models that considers effects of further number of factors and have less number of assumptions will bring significant benefits to studying the departure decisions of all students during post-secondary education. More accurate models will enable all decision makers such as policy makers, managers, and teachers to make informed decision, and provide better services. This in turn can increase educational levels, benefits the economy, and positively affects society.

APPENDIX A: UCF IRB LETTER

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

UNIVERSITY OF CENTRAL FLORIDA

NOT HUMAN RESEARCH DETERMINATION

February 21, 2020

Dear Marieh Alaghband:

On 2/21/2020, the IRB reviewed the following protocol:

Type of Review:	Initial Study
Title of Study:	FePh: An Annotated Facial Expression Dataset for the RWTH-PHOENIX-Weather 2014 Dataset
Investigator:	Marieh Alaghband
IRB ID:	STUDY00001501
Funding:	None
Grant ID:	None
Documents Reviewed:	• HRP-250 – Request for Not Human Subject Research determination form, Category: IRB Protocol; • List of the data points, variables, and/or information that will be collected and/or analyzed, Category: Letters of Support;

The IRB determined that the proposed activity is not research involving human subjects as defined by DHHS and FDA regulations.

IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities are research involving human in which the organization is engaged, please submit a new request to the IRB for a determination. You can create a modification by clicking **Create Modification / CR** within the study.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Adrienne Showman
Designated Reviewer

Page 1 of 1

APPENDIX B: PERMISSION LETTER FROM THE PUBLISHER

WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY
CONFERENCE REGISTRATION AND COPY RIGHT TRANSFER FORM
http://www.waset.org/

CONFERENCE PARCITIPANT'S INFORMATION

Full Name & Title (Prof; Dr; Ms; Mr)	Ivan Garibay Marie Alaghband will attend the conference		
Whether your attend the conference	Yes No		
Your Postal Address	4000 Central Florida Blvd Building ||, Room 321		
City: Orlando	State: Florida	United States	Postcode: 32816-2993
Telephone:	Fax:	Mobile: +1-310-362-6352	ivan.garibay@ucf.edu
Conference Date & Venue : Jan 21-22, 2021 London United Kingdom			
Conference Title : ICMI 2021 :23th Int. Conf. on Multimodal Interaction			
Paper(s) ID Code(s) (for Authors Only)	21UK010053		
Mode of Participation	Presenter	I	Listener
Paper Title : Facial Expression Phoenix: An Annotated Sequenced Dataset for Facial and Emotion-Specified Expressions in Sign Language			

(Please Compile This Form, Sign and Upload Through the Conference Website)

Please complete and sign this form and send it back to us with the final version of your manuscript. It is required to obtain a written confirmation from authors in order to acquire copyrights for papers published in World Academy of Science, Engineering and Technology.

Copyright Transfer Statement

The copyright to this article is transferred to World Academy of Science, Engineering and Technology if and when the article is accepted for publication. The undersigned hereby transfers any and all rights in and to the paper including without limitation all copyrights to World Academy of Science, Engineering and Technology. The undersigned hereby represents and warrants that the paper is original and that he/she is the author of the paper, except for material that is clearly identified as to its original source, with permission notices from the copyright owners where required. The undersigned represents that he/she has the power and authority to make and execute this assignment.

We declare that:

1. This paper has not been published in the same form elsewhere.
2. It will not be submitted anywhere else for publication prior to acceptance/rejection by Academy of Science, Engineering and Technology.
3. A copyright permission is obtained for materials published elsewhere and which require this permission for reproduction.

Furthermore, I/We hereby transfer the unlimited rights of publication of the above mentioned paper in whole to World Academy of Science, Engineering and Technology. The copyright transfer covers the exclusive right to reproduce and distribute the article, including reprints, translations, photographic reproductions, microform, electronic form (mobile, offline, online) or any other reproductions of similar nature.

The corresponding author signs for and accepts responsibility for releasing this material on behalf of any and all co-authors. This agreement is to be signed by at least one of the authors who has obtained the assent of the co-author(s) where applicable. After submission of this agreement signed by the corresponding author, changes of authorship or in the order of the authors listed will not be accepted.

Refund and Cancellation Policy : In case of cancellation, No refund of conference registration fees available after the conference registration fees are completely paid. Also, No refunds will be made for No Shows.

I hereby have read, understood and fully accepted the above mentioned Copyright Transfer Statement, Refund and Cancellation Policy conditions and Conference Programme as part of my conference registration.

Yours Sincerely,

Date 12/21/2020

Full Name Ivan Garibay

Signature Ivan Garibay

LIST OF REFERENCES

[1] V. Tinto, "Dropout from higher education: A theoretical synthesis of recent research," *Review of educational research*, vol. 45, no. 1, pp. 89–125, 1975.

[2] "World health organization, deafness and hearing loss, online, 2020. [online]. available: https://www.who.int/health-topics/hearing-loss." [Online]. Available: https://www.who.int/health-topics/hearing-loss

[3] H. R. V. Joze and O. Koller, "Ms-asl: A large-scale data set and benchmark for understanding american sign language," *arXiv preprint arXiv:1812.01053*, 2018.

[4] G. Matheson. Sign language alphabets from around the world. [Online]. Available: https://blog.ai-media.tv/blog/sign-language-alphabets-from-around-the-world#:~:text=There%20are%20over%20135%20different,Australian%20Sign%20Language%20(Auslan).

[5] what is american sign language? [Online]. Available: https://www.nad.org/

[6] O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: Enabling robust statistical continuous sign language recognition via hybrid cnn-hmms," *International Journal of Computer Vision*, vol. 126, no. 12, pp. 1311–1325, 2018.

[7] P. Kumar, P. P. Roy, and D. P. Dogra, "Independent bayesian classifier combination based sign language recognition using facial expression," *Information Sciences*, vol. 428, pp. 30–48, 2018.

[8] H.-D. Yang and S.-W. Lee, "Combination of manual and non-manual features for sign language recognition based on conditional random field and active appearance model," in *2011 International Conference on Machine Learning and Cybernetics*, vol. 4. IEEE, 2011, pp. 1726–1731.

[9] O. Koller, O. Zargaran, H. Ney, and R. Bowden, "Deep sign: hybrid cnn-hmm for continuous sign language recognition," in *Proceedings of the British Machine Vision Conference 2016*, 2016.

[10] D. H. Neiva and C. Zanchettin, "Gesture recognition: A review focusing on sign language in a mobile context," *Expert Systems with Applications*, 2018.

[11] D. Kelly, J. Reilly Delannoy, J. Mc Donald, and C. Markham, "A framework for continuous multimodal sign language recognition," in *Proceedings of the 2009 international conference on Multimodal interfaces*. ACM, 2009, pp. 351–358.

[12] A. Wadhawan and P. Kumar, "Sign language recognition systems: A decade systematic literature review," *Archives of Computational Methods in Engineering*, pp. 1–29, 2019.

[13] S. C. Agrawal, A. S. Jalal, and R. K. Tripathi, "A survey on manual and non-manual sign language recognition for isolated and continuous sign," *International Journal of Applied Pattern Recognition*, vol. 3, no. 2, pp. 99–134, 2016.

[14] M. W. Kadous *et al.*, *Temporal classification: Extending the classification paradigm to multivariate time series*. University of New South Wales Kensington, 2002.

[15] F. Ronchetti, F. Quiroga, C. A. Estrebou, L. C. Lanzarini, and A. Rosete, "Lsa64: an argentinian sign language dataset," in *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016).*, 2016.

[16] F. A. Freitas, S. M. Peres, C. A. Lima, and F. V. Barbosa, "Grammatical facial expression recognition in sign language discourse: a study at the syntax level," *Information Systems Frontiers*, vol. 19, no. 6, pp. 1243–1259, 2017.

[17] N. Pugeault and R. Bowden, “Spelling it out: Real-time asl fingerspelling recognition,” in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1114–1119.

[18] Z. Ren, J. Yuan, and Z. Zhang, “Robust hand gesture recognition based on finger-earth mover’s distance with a commodity depth camera,” in *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 2011, pp. 1093–1096.

[19] Z. A. Ansari and G. Harit, “Nearest neighbour classification of indian sign language gestures using kinect camera,” *Sadhana*, vol. 41, no. 2, pp. 161–182, 2016.

[20] O. Koller, J. Forster, and H. Ney, “Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers,” *Computer Vision and Image Understanding*, vol. 141, pp. 108–125, 2015.

[21] J. Forster, C. Schmidt, T. Hoyoux, O. Koller, U. Zelle, J. H. Piater, and H. Ney, “Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus.” in *LREC*, 2012, pp. 3785–3789.

[22] A. M. Martínez, R. B. Wilbur, R. Shay, and A. C. Kak, “Purdue rvl-slll asl database for automatic recognition of american sign language,” in *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*. IEEE, 2002, pp. 167–172.

[23] N. K. Caselli, Z. S. Sehyr, A. M. Cohen-Goldberg, and K. Emmorey, “Asl-lex: A lexical database of american sign language,” *Behavior research methods*, vol. 49, no. 2, pp. 784–801, 2017.

[24] M. Alaghband, N. Yousefi, and I. Garibay, “Feph: An annotated facial expression dataset for the rwth-phoenix-weather 2014 dataset,” *arXiv preprint arXiv:2003.08759*, 2020.

[25] S. Escalera, X. Baró, J. Gonzalez, M. A. Bautista, M. Madadi, M. Reyes, V. Ponce-López, H. J. Escalante, J. Shotton, and I. Guyon, "Chalearn looking at people challenge 2014: Dataset and results," in *Workshop at the European Conference on Computer Vision*. Springer, 2014, pp. 459–473.

[26] C. N. Nyaga and R. D. Wario, "Sign language gesture recognition through computer vision," in *2018 IST-Africa Week Conference (IST-Africa)*. IEEE, 2018, pp. Page–1.

[27] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955," *AI magazine*, vol. 27, no. 4, pp. 12–12, 2006.

[28] I. B. A. TURING, "Computing machinery and intelligence-am turing," *Mind*, vol. 59, no. 236, p. 433, 1950.

[29] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315–323.

[30] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters*, vol. 42, pp. 11–24, 2014.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[32] Y. Bengio, A. C. Courville, and P. Vincent, "Unsupervised feature learning and deep learning: A review and new perspectives," *CoRR, abs/1206.5538*, vol. 1, p. 2012, 2012.

[33] Y. Bengio, Y. LeCun *et al.*, "Scaling learning algorithms towards ai," *Large-scale kernel machines*, vol. 34, no. 5, pp. 1–41, 2007.

[34] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, “Why does unsupervised pre-training help deep learning?” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 201–208.

[35] L. Deng and D. Yu, “Deep learning: methods and applications,” *Foundations and trends in signal processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[36] L. Deng, “Three classes of deep learning architectures and their applications: a tutorial survey,” *APSIPA transactions on signal and information processing*, 2012.

[37] R. Salakhutdinov and G. Hinton, “Deep boltzmann machines,” in *Artificial intelligence and statistics*, 2009, pp. 448–455.

[38] I. Arel, D. C. Rose, and T. P. Karnowski, “Deep machine learning-a new frontier in artificial intelligence research [research frontier],” *IEEE computational intelligence magazine*, vol. 5, no. 4, pp. 13–18, 2010.

[39] D. L. Boutin, “Persistence in postsecondary environments of students with hearing impairments,” *Journal of Rehabilitation*, vol. 74, no. 1, p. 25, 2008. [Online]. Available: https://search.proquest.com/docview/236328417?accountid=10003

[40] O. Aljohani, “A comprehensive review of the major studies and theoretical models of student retention in higher education.” *Higher education studies*, vol. 6, no. 2, pp. 1–18, 2016.

[41] National Council on Disability (US), *People with Disabilities and Postsecondary Education: Position Paper*. National Council on Disability, 2003. [Online]. Available: https://ncd.gov/publications/2003/people-disabilities-and-postsecondary-education-position-paper

[42] K. Mamiseishvili and L. C. Koch, "First-to-second-year persistence of students with disabilities in postsecondary institutions in the united states," *Rehabilitation Counseling Bulletin*, vol. 54, no. 2, pp. 93–105, 2011.

[43] J. L. Luckner and J. Stewart, "Self-assessments and other perceptions of successful adults who are deaf: An initial investigation," *American Annals of the Deaf*, vol. 148, no. 3, pp. 243–250, 2003.

[44] C. L. Garberoglio, S. Schoffstall, S. Cawthon, M. Bond, and J. Ge, "The role of self-beliefs in predicting postschool outcomes for deaf young adults," *Journal of Developmental and Physical Disabilities*, vol. 26, no. 6, pp. 667–688, 2014.

[45] D. W. Test, V. L. Mazzotti, A. L. Mustian, C. H. Fowler, L. Kortering, and P. Kohler, "Evidence-based secondary transition predictors for improving postschool outcomes for students with disabilities," *Career Development for Exceptional Individuals*, vol. 32, no. 3, pp. 160–181, 2009.

[46] C. Garberoglio, J. Palmer, and S. Cawthon, "Undergraduate enrollment of deaf students in the united states," 2019. [Online]. Available: https://www.nationaldeafcenter.org/resource/undergraduate-enrollment-deaf-students-united-states

[47] C. Garberoglio, J. Palmer, S. Cawthon, and A. Sales, "Deaf people and educational attainment in the united states: 2019," *Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Deaf Center on Postsecondary Outcomes*, 2019a. [Online]. Available: https://www.nationaldeafcenter.org/resource/deaf-people-and-educational-attainment-united-states

[48] C.-C. Wang and K.-C. Wang, "Hand posture recognition using adaboost with sift for human robot interaction," in *Recent progress in robotics: viable robotic service to human*. Springer, 2007, pp. 317–329.

[49] S. F. Chevtchenko, R. F. Vale, and V. Macario, "Multi-objective optimization for hand posture recognition," *Expert Systems with Applications*, vol. 92, pp. 170–181, 2018.

[50] H. Birk, T. B. Moeslund, and C. B. Madsen, "Real-time recognition of hand alphabet gestures using principal component analysis," in *Proceedings of the Scandinavian conference on image analysis*, vol. 1. PROCEEDINGS PUBLISHED BY VARIOUS PUBLISHERS, 1997, pp. 261–268.

[51] H. N. Saha, S. Tapadar, S. Ray, S. K. Chatterjee, and S. Saha, "A machine learning based approach for hand gesture recognition using distinctive feature extraction," in *Computing and Communication Workshop and Conference (CCWC), 2018 IEEE 8th Annual*. IEEE, 2018, pp. 91–98.

[52] H. Luqman, S. A. Mahmoud *et al.*, "Transform-based arabic sign language recognition," *Procedia Computer Science*, vol. 117, pp. 2–9, 2017.

[53] W. Tao, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using convolutional neural networks with multiview augmentation and inference fusion," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 202–213, 2018.

[54] M. Beena, M. A. Namboodiri, and P. Dean, "Automatic sign language finger spelling using convolution neural network: Analysis," *International Journal of Pure and Applied Mathematics*, vol. 117, no. 20, pp. 9–15, 2017.

[55] B. Hisham and A. Hamouda, "Arabic static and dynamic gestures recognition using leap motion," *J. Comput. Sci*, vol. 13, no. 8, pp. 337–354, 2017.

[56] J. Huang, W. Zhou, H. Li, and W. Li, "Sign language recognition using 3d convolutional neural networks," in *Multimedia and Expo (ICME), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1–6.

[57] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, “Coupled hmm-based multi-sensor data fusion for sign language recognition,” *Pattern Recognition Letters*, vol. 86, pp. 1–8, 2017.

[58] ——, “A multimodal framework for sensor based sign language recognition,” *Neurocomputing*, vol. 259, pp. 21–38, 2017.

[59] I. Rasines, A. Remazeilles, and P. M. I. Bengoa, “Feature selection for hand pose recognition in human-robot object exchange scenario,” in *Emerging Technology and Factory Automation (ETFA), 2014 IEEE*. IEEE, 2014, pp. 1–8.

[60] A. Barczak, N. Reyes, M. Abastillas, A. Piccio, and T. Susnjak, “A new 2d static hand gesture colour image dataset for asl gestures,” 2011.

[61] S. Masood, H. C. Thuwal, and A. Srivastava, “American sign language character recognition using convolution neural network,” in *Smart Computing and Informatics*. Springer, 2018, pp. 403–412.

[62] C. Ong, I. Lim, J. Lu, C. Ng, and T. Ong, “Sign-language recognition through gesture & movement analysis (sigma),” in *Mechatronics and Machine Vision in Practice 3*. Springer, 2018, pp. 235–245.

[63] N. M. Kakoty and M. D. Sharma, “Recognition of sign language alphabets and numbers based on hand kinematics using a data glove,” *Procedia Computer Science*, vol. 133, pp. 55–62, 2018.

[64] W. Kadous *et al.*, “Grasp: Recognition of australian sign language using instrumented gloves,” 1995.

[65] S. Masood, A. Srivastava, H. C. Thuwal, and M. Ahmad, “Real-time sign language gesture (word) recognition from video sequences using cnn and rnn,” in *Intelligent Engineering Informatics*. Springer, 2018, pp. 623–632.

[66] O. Koller, H. Ney, and R. Bowden, “Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3793–3802.

[67] N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, “Subunets: End-to-end hand shape and continuous sign language recognition,” in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017, pp. 3075–3084.

[68] R. Cui, H. Liu, and C. Zhang, “Recurrent convolutional neural networks for continuous sign language recognition by staged optimization,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7361–7369.

[69] O. Koller, S. Zargaran, and H. Ney, “Re-sign: Re-aligned end-to-end sequence modelling with deep recurrent cnn-hmms,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4297–4305.

[70] R. Cui, H. Liu, and C. Zhang, “A deep neural framework for continuous sign language recognition by iterative training,” *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1880–1891, 2019.

[71] B. P. Kumar and M. Manjunatha, “A hybrid gesture recognition method for american sign language,” *Indian Journal of Science and Technology*, vol. 10, no. 1, 2017.

[72] M. Tolba and A. Elons, “Recent developments in sign language recognition systems,” in *Computer Engineering & Systems (ICCES), 2013 8th International Conference on*. IEEE, 2013, pp. xxxvi–xlii.

[73] C. Darwin and P. Prodger, *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.

[74] Y. Li, X. Wang, W. Liu, and B. Feng, “Deep attention network for joint hand gesture localization and recognition using static rgb-d images,” *Information Sciences*, vol. 441, pp. 66–78, 2018.

[75] Y. Fan, J. C. Lam, and V. O. Li, “Multi-region ensemble convolutional neural network for facial expression recognition,” in *International Conference on Artificial Neural Networks*. Springer, 2018, pp. 84–94.

[76] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, “Hybrid deep neural networks for face emotion recognition,” *Pattern Recognition Letters*, vol. 115, pp. 101–106, 2018.

[77] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, “Training deep networks for facial expression recognition with crowd-sourced label distribution,” in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 279–283.

[78] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, “Joint fine-tuning in deep neural networks for facial expression recognition,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2983–2991.

[79] I. M. Revina and W. S. Emmanuel, “A survey on human face expression recognition techniques,” *Journal of King Saud University-Computer and Information Sciences*, 2018.

[80] X. Zheng, Y. Guo, H. Huang, Y. Li, and R. He, “A survey of deep facial attribute analysis,” *International Journal of Computer Vision*, pp. 1–33, 2020.

[81] H. Zhou, W. Zhou, Y. Zhou, and H. Li, “Spatial-temporal multi-cue network for continuous sign language recognition.” in *AAAI*, 2020, pp. 13 009–13 016.

[82] H.-D. Yang and S.-W. Lee, “Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine,” *Pattern Recognition Letters*, vol. 34, no. 16, pp. 2051–2056, 2013.

[83] S. Chen, Y. Tian, Q. Liu, and D. N. Metaxas, “Recognizing expressions from face and body gesture by temporal normalized motion and appearance features,” *Image and Vision Computing*, vol. 31, no. 2, pp. 175–185, 2013.

[84] J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, “Video-based sign language recognition without temporal segmentation,” *arXiv preprint arXiv:1801.10111*, 2018.

[85] L. Quesada, G. Marín, and L. A. Guerrero, “Sign language recognition model combining non-manual markers and handshapes,” in *International Conference on Ubiquitous Computing and Ambient Intelligence*. Springer, 2016, pp. 400–405.

[86] C. M. Jin, Z. Omar, and M. H. Jaward, “A mobile application of american sign language translation via image processing algorithms,” in *Region 10 symposium (tensymp), 2016 ieee*. IEEE, 2016, pp. 104–109.

[87] K. S. Warrier, J. K. Sahu, H. Halder, R. Koradiya, and V. K. Raj, “Software based sign language converter,” in *Communication and Signal Processing (ICCSP), 2016 International Conference on*. IEEE, 2016, pp. 1777–1780.

[88] P. Loke, J. Paranjpe, S. Bhabal, and K. Kanere, “Indian sign language converter system using an android app,” in *Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of*, vol. 2. IEEE, 2017, pp. 436–439.

[89] M. S. G. Shinde, R. R. Itkarkar, and A. V. Nandi, “Gesture to speech conversion for sign language recognition,” 2017.

[90] O. H. Al-Barahamtoshy and H. M. Al-Barhamtoshy, “Arabic text-to-sign (artts) model from automatic sr system,” *Procedia Computer Science*, vol. 117, pp. 304–311, 2017.

[91] A. Kurakin, Z. Zhang, and Z. Liu, “A real time system for dynamic hand gesture recognition with a depth sensor.” in *EUSIPCO*, vol. 2, no. 5, 2012, p. 6.

[92] J. Wan, Y. Zhao, S. Zhou, I. Guyon, S. Escalera, and S. Z. Li, “Chalearn looking at people rgb-d isolated and continuous datasets for gesture recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 56–64.

[93] T. Kapuscinski, M. Oszust, M. Wysocki, and D. Warchol, “Recognition of hand gestures observed by depth cameras,” *International Journal of Advanced Robotic Systems*, vol. 12, no. 4, p. 36, 2015.

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

[95] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, “American sign language recognition with the kinect,” in *Proceedings of the 13th international conference on multimodal interfaces*. ACM, 2011, pp. 279–286.

[96] D. Uebersax, J. Gall, M. Van den Bergh, and L. Van Gool, “Real-time sign language letter and word recognition from depth data,” in *2011 IEEE international conference on computer vision workshops (ICCV Workshops)*. IEEE, 2011, pp. 383–390.

[97] N. Cihan Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, “Neural sign language translation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7784–7793.

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

[99] B. Feng, F. He, X. Wang, Y. Wu, H. Wang, S. Yi, and W. Liu, "Depth-projection-map-based bag of contour fragments for robust hand gesture recognition," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 511–523, 2017.

[100] R. Wilbur and A. C. Kak, "Purdue rvl-slll american sign language database," 2006.

[101] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, Q. Yuan, and A. Thangali, "The american sign language lexicon video dataset," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*. IEEE, 2008, pp. 1–8.

[102] T. Shanableh, K. Assaleh, and M. Al-Rousan, "Spatio-temporal feature-extraction techniques for isolated gesture recognition in arabic sign language," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 3, pp. 641–650, 2007.

[103] J. Forster, C. Schmidt, O. Koller, M. Bellgardt, and H. Ney, "Extensions of the sign language recognition and translation corpus rwth-phoenix-weather." in *LREC*, 2014, pp. 1911–1916.

[104] U. Von Agris, M. Knorr, and K.-F. Kraiss, "The significance of facial features for automatic sign language recognition," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–6.

[105] M. Oliveira, H. Chatbri, Y. Ferstl, M. Farouk, S. Little, N. E. O'Connor, and A. Sutherland, "A dataset for irish sign language recognition," 2017.

[106] H. Hosoe, S. Sako, and B. Kwolek, “Recognition of jsl finger spelling using convolutional neural networks,” in *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*. IEEE, 2017, pp. 85–88.

[107] M. Stinson and G. Walter, “Improving retention for deaf and hard of hearing students: What the research tells us,” *JADARA-ROCHESTER NY-*, vol. 30, pp. 14–23, 1997.

[108] M. Wagner, T. W. Cadwallader, C. Marder, R. Cameto, D. Cardoso, N. Garza, P. Levine, and L. Newman, “Life outside the classroom for youth with disabilities,” *Menlo Park, CA: SRI International*, 2003. [Online]. Available: https://nlts2.sri.com/reports/2003_04-2/nlts2_report_2003_04-2_complete.pdf

[109] National Deaf Center on Postsecondary Outcomes, “Non-cognitive factors that support postsecondary persistence in deaf students,” 2019, accessed: 5-3-2019. [Online]. Available: https://www.nationaldeafcenter.org/file/non-cognitive-factors-support-postsecondary-persistence-deaf-students

[110] L. A. Newman, M. Marschark, D. M. Shaver, and H. Javitz, “Course-taking effect on postsecondary enrollment of deaf and hard of hearing students,” *Exceptionality*, vol. 25, no. 3, pp. 170–185, 2017.

[111] R. H. Harrison, “Identifying factors that predict deaf students’ academic success in college,” Ph.D. dissertation, Graduate School of Arts and Sciences, University of Pennsylvania, 1987.

[112] M. S. Stinson, M. J. Scherer, and G. G. Walter, “Factors affecting persistence of deaf college students,” *Research in Higher Education*, vol. 27, no. 3, pp. 244–258, 1987.

[113] J. A. Cuculick and R. R. Kelly, “Relating deaf students’ reading and language scores at college entry to their degree completion rates,” *American annals of the deaf*, vol. 148, no. 4, pp. 279–286, 2003.

[114] J. T. Richardson, J. MacLeod-Gallinger, B. G. McKee, and G. L. Long, "Approaches to studying in deaf and hearing students in higher education," *Journal of Deaf Studies and Deaf Education*, vol. 5, no. 2, pp. 156–173, 2000.

[115] M. S. Stinson and G. G. Walter, "Persistence in college," *Deaf students in postsecondary education*, pp. 43–64, 1992.

[116] C. L. Garberoglio, S. W. Cawthon, and M. Bond, "Assessing english literacy as a predictor of postschool outcomes in the lives of deaf individuals," *Journal of deaf studies and deaf education*, vol. 19, no. 1, pp. 50–67, 2014.

[117] J. A. Albertini, R. R. Kelly, and M. K. Matchett, "Personal factors that influence deaf college students' academic success," *Journal of deaf studies and deaf education*, vol. 17, no. 1, pp. 85–101, 2012.

[118] B. Danermark, "Persistence and academic and social integration of hearing-impaired students in postsecondary education: A review of research," *JADARA*, vol. 29, no. 2, p. 8, 1995. [Online]. Available: https://repository.wcsu.edu/jadara/vol29/iss2/8

[119] S. W. Cawthon, J. M. Caemmerer, D. M. Dickson, O. L. Ocuto, J. Ge, and M. P. Bond, "Social skills as a predictor of postsecondary outcomes for individuals who are deaf," *Applied Developmental Science*, vol. 19, no. 1, pp. 19–30, 2015.

[120] S. Rogers, K. Muir, and C. R. Evenson, "Signs of resilience: Assets that support deaf adults' success in bridging the deaf and hearing worlds," *American Annals of the Deaf*, vol. 148, no. 3, pp. 222–232, 2003.

[121] N. Gilbert, *Agent-based models*. Sage Publications, Incorporated, 2019, vol. 153. [Online]. Available: http://epubs.surrey.ac.uk/1580/1/fulltext.pdf

[122] N. Yousefi, M. Alaghband, and I. Garibay, "A comprehensive survey on machine learning techniques and user authentication approaches for credit card fraud detection," *arXiv preprint arXiv:1912.02629*, 2019.

[123] U. Wilensky and W. Rand, *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Mit Press, 2015.

[124] J. M. Epstein, "Agent-based computational models and generative social science," *Complexity*, vol. 4, no. 5, pp. 41–60, 1999.

[125] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the national academy of sciences*, vol. 99, no. suppl 3, pp. 7280–7287, 2002.

[126] "Labelbox, "labelbox," online, 2019. [online]. available: https://labelbox.com." [Online]. Available: https://labelbox.com/

[127] V. Tinto, "Leaving college rethinking the causes and cures of student attrition.: The university of chicago," *Chicago, IL*, 1993.

[128] D. Kember, *Open learning courses for adults: A model of student progress*. Educational Technology, 1995.

[129] U. Wilensky *et al.*, "Center for connected learning and computer-based modeling," in *NetLogo*. Northwestern University, 1999.