

Hand Gesture Vocalizer for Dumb and Deaf People

Sanish Manandhar¹, Sushana Bajracharya², Sanjeev Karki³, Ashish Kumar Jha⁴

^[1,2,3]Department of Computer Science, Nepal Engineering College, Changunarayan, Bhaktapur, Nepal

¹sanish.manandhar20@gmail.com, ²sushanabajra96@gmail.com, ³sanjeevkarki77@gmail.com

⁴Assistant Professor, ashishkj@nec.edu.np



Sanish Manandhar is a student of BE computer Engineering, Batch 2014 at Nepal Engineering College and recently he has received his bachelor degree from Pokhara University. He is

currently working as web developer in a prominent software company.



Sushana Bajracharya is a student of BE computer Engineering, Batch 2014 at Nepal Engineering College and recently she has received her bachelor degree from Pokhara University. She is currently working as web

developer in a prominent software company.



Sanjeev Karki is a student of BE computer Engineering, Batch 2014 at Nepal Engineering College and recently he has received his bachelor degree from Pokhara University. He is currently working

as web developer in a prominent software company.



Ashish Kumar Jha received his Master's degree in Computer Science specialization in Networking from Sharda University in 2017. He has involved in software development and teaching profession since

2013, currently working as assistant professor in Nepal Engineering College. His research interest includes Internet of things, Image Processing and Pattern Recognition.

Abstract:

The main purpose of this paper is to confer the system that converts a given sign used by disabled person into its appropriate textual, audio, and pictorial form using components such as Arduino Mega, Flex sensors, Accelerometer, which could be understand by a common person. A wearable glove controller is design with flex sensors attached on each finger, which allows the system to sense the finger movements, and a Gy-61 accelerometer, which are uses to sense the hand movement of the disabled person. The wearable input glove controller sends the collected input signal to the system for processing. The system uses Random forest algorithm to predict the correct output to an accuracy of 85% on current training model.

Keywords: *Random Forest, Speech synthesis, Flex sensors, Data glove, Smart Glove.*

I. Introduction

Humans possess the voice capability for interaction and communication among each other. Unfortunately, not everybody has the capability of speaking and hearing. Sign language used among the community of people who cannot speak or hear as the means of communication. Sign language is a gesture representation that involves simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to express fluently with a speaker's thoughts. The people who cannot speak makes use of the sign languages to communicate with other fellow vocally impaired person and even with other normal people who knows the meanings of sign languages or an interpreter is needed to translate the meanings of sign languages to other people who can speak and do not know the meanings of sign languages. However, it is not always possible for an individual to be around all the time to interpret the sign languages

and not everybody can learn the sign languages. Thus, another alternative is that we can use a computer or a smart phone as a mediator. The computer or a smart phone could take an input from the vocally impaired person and give its textual as well as and audio form of output.

II. Related Work

A Lexicon Statics Survey of Signed Language in Nepal This thesis proposed by Hope M. Hurl but after his field study about the variety of sign languages used in Nepal and its similarity or dissimilarity with the American Sign Language (ASL) and British Sign Language (BSL) through the collection of world list and interviews. The results of the world list comparisons showed that there are at least three signed languages in the country [1]. In more recent years NpSL has been introduced to many Hearing Impaired Associations have undertaken to open classes for hearing impaired children in many places where the government does not yet have a school for the hearing impaired. In this course, a research conducted in the following cities of Nepal – Kathmandu, Surkhet, Jumla, Pokhara, Ghandruk, Dharan, and Rajbiraj where it was discovered that there was one national sign language used by all hearing impaired named the NpSL. The survey further concluded that, there were 3 village sign language that were used in the Indian Sign Language (ISL), ASL and BSL and the most widely used technique was the sign writing which is somewhat similar to imitate the object used in spoken language.

In this area, there has been a lot work done with implementation different sign language and technologies few work are discuss below:

A. Kinetic sign language translator [2]

This project was a result of collaboration, facilitated by Microsoft Research, between the Chinese Academy of Sciences, Beijing Union University, and Microsoft Research Asia, in which all the organization has made crucial contributions. The system understands the gestures of sign language and converts them to spoken and written language and vice versa. This system captures a conversation from both sides, displays the signer and renders a written and spoken translation of the sign language in real-time. It also takes the spoken words of non-signer and turns them into accurate, understandable sign language.

B. Sign Language Recognition System [3] [4]

Yang Quan, a Chinese student, defined a Basic Sign Language Recognition system that is able to translate a sequence of signs into the commonly used speech language and vice versa. The sign language/speech bidirectional translation (from signs to speech and speech to signs) focused on the Chinese Manual Alphabet where every single sign belongs to a single letter from the alphabet. The system was composed of a camera, a video display terminal (VDT), a speaker, a microphone, and a keyboard. Two kinds of data were used which were; vector of hand gestures and vector of lip actions. In order to characterize these vectors, they used the Normalized Moment of Inertia (NMI) algorithm and Hu moments. The former attempts to solve translation invariability, size invariability, anti-gradation distortion, and so on and the latter is a set of algebraic invariants that combines regular moments and is useful to describe the information of the image. They do not vary under change of size, translation, and rotation and have been widely used in Pattern Recognition and successfully proven in sign language letter recognition. As said before, they combine the hand gestures recognition with the lips movement reader in order to make the system more accurate. By using a multi-features SVMs classifier trained with a linear kernel, the 30 letters from the Chinese manual alphabet recognized with average accuracy of 95.55%.

C. Microcontroller and Sensors Based Gesture Vocalizer [5]

Sensors Based Gesture Vocalizer describes the design and working of a system that is useful for dumb, deaf and blind people to communicate with each another and with the normal people as well. Gesture Vocalizer is a large-scale multi-microcontroller based system that designed to facilitate the communication among the dumb, deaf and blind communities with the normal people. In the project a data glove is used which can detect almost all the movements of a hand and microcontroller based system converts some specified movements into human recognizable voice. The data glove is equipped with two types of sensors, which are; the bend sensors and accelerometers as tilt sensors. This system is beneficial for dumb people and their hands will

speak having worn the gesture, vocalizer data glove. This project is similar to our proposed project but it does not make use of any smart phone, instead it uses external speakers and a LCD display of outputs.

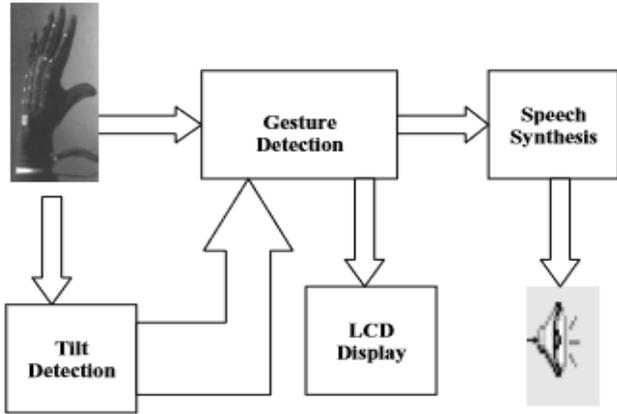


Figure 1: Block diagram of Microcontroller and Sensors Based Gesture Vocaliser

D. Hand Gesture Recognition Using KINECT [6]

This is an automatic Gesture Recognition System, which uses Microsoft Kinect for capturing the raw image of the gesture. Originally developed for gaming, the Kinect's sensors read a user's body position and movements and, with the help of a computer, Gesture Recognition System translates them into words. This system thus has tremendous potential for understanding the complex gestures and for translating the signs into spoken or written words. It check textual character in train data and display as output if present. If no match found, then the system allows the user to store the new gesture and its corresponding textual form. The system allows the user to auto-create a dictionary. The main motive of this project is to make interaction possible between an unimpaired and hearing-impaired individual.

III. Hardware Details: Flex Sensors

The flex sensors also referred as the bend sensor is a type of sensor that measures the amount of bending or deflection or flexing as the name itself has it. As the sensors are bent the sensors provides us with an electrical resistance value, the sensor gives so the more the sensors are bent, more resistance value. When measured with a multi meter, the flex sensor

at a flat and steady position gives a resistance value near to 25k ohm whereas when the sensors are fully bent they give a resistance value near to 72k ohm. They are usually in the form of a thin strip, are very comfortable to use, as they are very light in weight, and are easily bent [8].

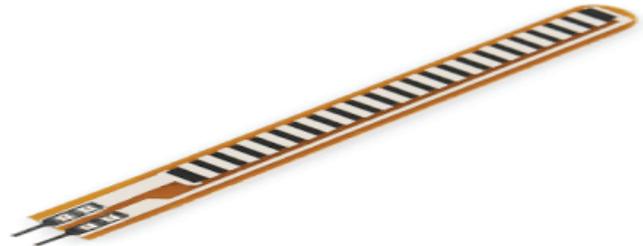


Figure 2: Flex sensor

One side of the sensor is printed with a polymer ink that has conductive particles embedded on it, and note that the flex sensors gives a proper change in resistance value only when it is bent away from the ink like and provides very minimum change in resistance value when it is bent in the reverse direction.

The flex sensor comes in various sizes and made by various companies, and in our research, we have used 10 flex sensors of 2.2 inches each placed on five fingers. Each flex sensor consists of two strips, one is a thin strip connected to the high voltage (that can be 5v or 3.3v) and another is a fat strip consisting of the polymer ink connected to ground through a resistor and provides the resistance output as well.

By connecting the flex sensor with a static resistor to create a voltage divider, it produces a variable voltage that can read by a microcontroller's analog-to-digital converter [9].

The accelerometer can use to measure the static acceleration of gravity in tilt- sensing applications and can used to measure dynamic acceleration resulting from motion, shock, or vibrations. By measuring the amount of static acceleration due to gravity, we can find out the angle the device tilted at with respect to the earth and by sensing the amount of dynamic acceleration, you can analyze the way the device is moving. There are many different ways to make an accelerometer.

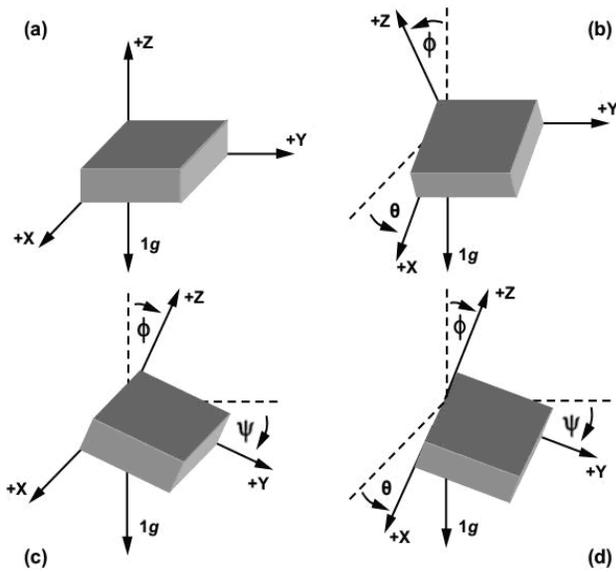


Figure 3: Tilt and inclination of the accelerometer

Acceleration along the axes deflects the movable mass. This displacement of moving plate (mass) unbalances the differential capacitor, which results in sensor output. Output amplitude is proportional to acceleration. 16-bit ADC used to get digitized output. The full-scale range of acceleration is +/- 2g, +/- 4g, +/- 8g, +/- 16g. It measures in g (gravity force) unit. When device placed on flat surface it will measure 0g on X and Y-axis and +1g on Z-axis [11].

IV. Methodologies

The classification of this article is based on Random Forest Algorithm.

A. Random forest algorithm

Random forests or random decision forests are an ensemble learning method or supervised classification algorithm for classification, regression and other tasks, that operate by constructing a large number of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. This algorithm creates a forest with a number of trees and the more trees in the forest the more robust the forest looks like.

Random decision forests correct the decision trees' habit of over fitting to their training set. Ensemble algorithms are those, which combine more than one algorithms of same or different kind for

classifying objects [13]. For example, running a prediction over Naive Bayes, SVM and Decision Tree and then taking vote for final consideration of class for test object.

Similarly, in the random forest classifier, higher number of trees in the forest results in higher accuracy of the outputs. Random forest classifier creates a set of decision trees from randomly selected subset of training set and then it aggregates the votes from different decision trees to decide the final class of the test object.

Suppose training set is given as: $[X_1, X_2, X_3, X_4]$ with corresponding labels as $[L_1, L_2, L_3, L_4]$, random forest may create three decision trees taking input of subset for example [14], $[X_1, X_2, X_3]$, $[X_1, X_2, X_4]$ & $[X_2, X_3, X_4]$

So finally, it makes predictions based on the majority of votes made from each of the decision trees. This algorithm works well because a single decision tree may be prone to a noise, but aggregation of many decision trees reduce the effect of noise gave more accurate result for a different sets of input. The random forest can apply weight concept for considering the impact of result from any decision tree. Tree with high error rate give low weight value and vice versa. There are two stages in this algorithm, one is random forest creation, and the other is to make a prediction from the random forest classifier created in the first stage.

The whole process is explained below.

Random Forest creation pseudo code:

1. Randomly select "k" features from total "m" features.
 - a. Where $k \ll m$
2. Among the "k" features, calculate the node "d" using the best split point.
3. Split the node into daughter nodes using the best split.
4. Repeat 1 to 3 steps until "l" number of nodes has reached.
5. Build forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.

Random forest prediction pseudo-code:

- Takes the test features, use the rules of each randomly created decision tree to predict the outcome, and stores the predicted outcome (target).
- Calculate the votes for each predicted target.
- Consider the high voted predicted target as the final prediction from the random forest algorithm.

The over fitting problem will never come when we use the random forest algorithm in any classification problem.

- The same random forest algorithm can be used for both classification and regression task.
- The random forest algorithm could be used for feature engineering.
- It means identifying the most important features out of the available features from the training dataset.

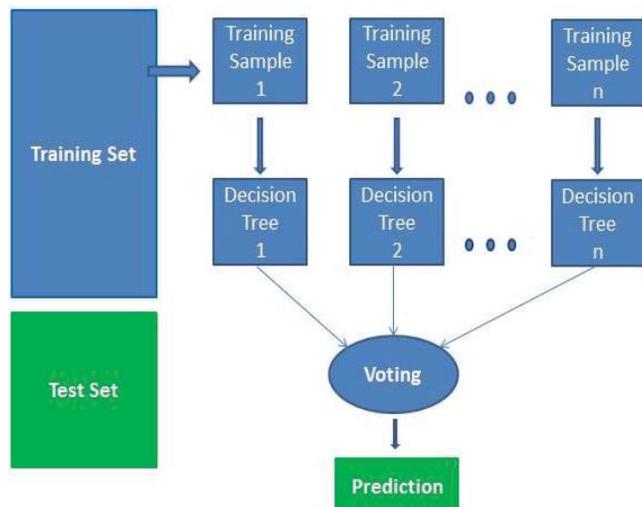


Figure 5: Working of Random Forest algorithm [15]

Random forest considered as a highly accurate and robust method. It does not suffer from the over fitting problem, this algorithm can be used in both classification and regression problems.

Random forest can also handle missing values, with two ways: using median values to replace continuous variables, and computing the proximity- weighted average of missing values.

Random forest is slow in generating predictions because it has multiple decision trees. This algorithm

is time – consuming. The model is difficult to interpret compared to a decision tree, where you can easily make a decision by following the path in the tree.

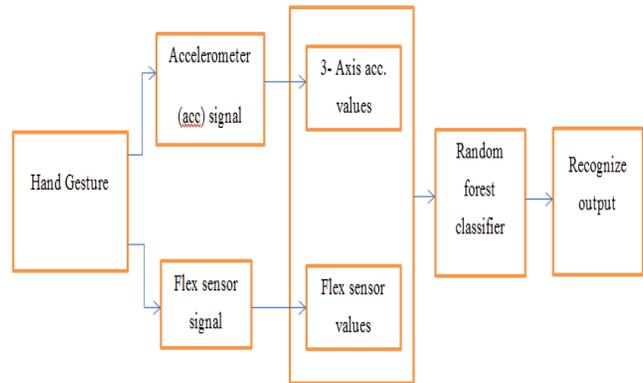


Figure 6: Feature Extraction Diagram

When a user makes a hand gesture, eight user inputs from the glove controller has given to the system in which there is 3- axes accelerometer signal and five flex sensor signals. Three axes values and five flex sensor values that has taken from the user acts as the features for recognizing a particular gesture. These features for the gesture has passed through random forest classifier, then the gestures has classified according to the features, and the output is recognized.

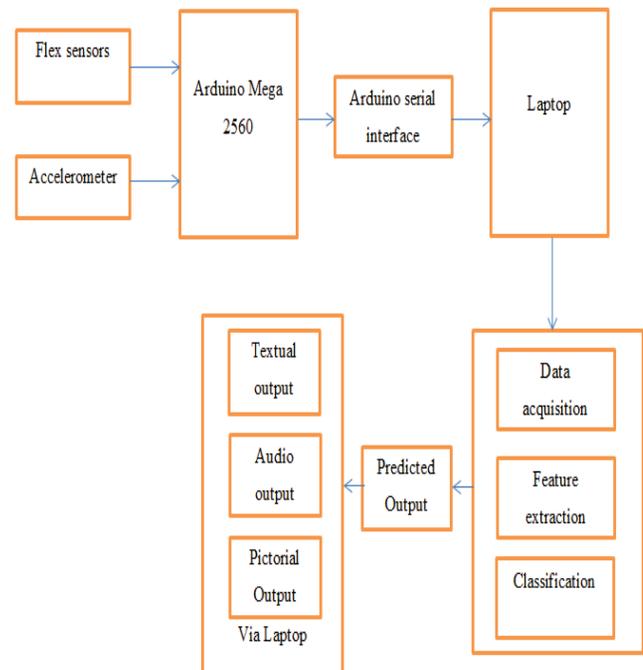


Figure 7: System Block Diagram

With the machine, has trained with those dataset, for the real time application of the trained machine, for a particular gesture the flex sensor and accelerometer data were collected from the glove movement and processed by the Arduino converting those raw data to meaningful data. These values help in feature extraction process. Those data has sent from the Arduino to the serial port, has collected by the python through the serial interface and saved as temporary dataset (CSV format) form. Finally, the temporary dataset passed through the machine such that the machine extracts the features of those dataset and predicts the appropriate output for that gesture with reference to the Random Forest Classifier model. As the output has recognized by the model, laptop screen and speaker has used to display the output to the user.

V. Algorithms

A. Algorithm for dataset preparation at Arduino Mega 2560

Step 1: Start

Step 2: Input: For a particular gesture, 8 sensor data inputs (5 flex sensors data and 3 accelerometer data)

Step 3: Convert raw sensor data to meaningful data

Step 4: Set the baud rate

Step 5: Send data to the serial port

Step 6: Is Arduino buffer flushed?

 If yes, go to step 7

 Else, wait for buffer to flush then continue

Step 7: Add delay of 10 milliseconds.

Step 8: Go to step 2.

B. Algorithm for real time application

Step 1: Start

Step 2: Train the model using the training dataset.

Step 3: Set the baud rate similar to Arduino serial baud rate and open the serial connection.

Step 4: Is there data at serial port?

 If no, wait till data available then continue.

 Else, go to step 5.

Step 5: Input: Collect data from the serial port at real time.

Step 6: Open new file in write mode.

Step 7: Write the collected data the new file in CSV format.

Step 8: Flush the Arduino buffer.

Step 9: Close the Serial port connection.

Step 9: Pass the temporary dataset through the model.

Step 10: Trained machine predicts the output.

Step 11: Display the prediction on the screen and play the audio of the predicted word.

Step 12: Add 1-second delay

Step 13: End

VI. Results

Finally, we have obtained a system that can read the values for a particular gesture done by the user, predict the output for the gesture, display it on the laptop screen, and provide an audio output via a desktop GUI. The dataset for all the alphabets and frequently used words has created. Using those datasets, the machine has trained and the model created. Then dataset for all the alphabets combined and shuffled in order to train the machine with reduced variance and make sure that model remains general and over fit less. When the input has taken for sign language of the alphabet or word, the input passed through the trained machine and the closest value predicted display and the audio played as output. Then the correlation plot of each alphabet was compare with other alphabet to check the level of closeness of alphabets with each other and predict the accuracy of the model. The accuracy for our model founded to be 96.8 %.

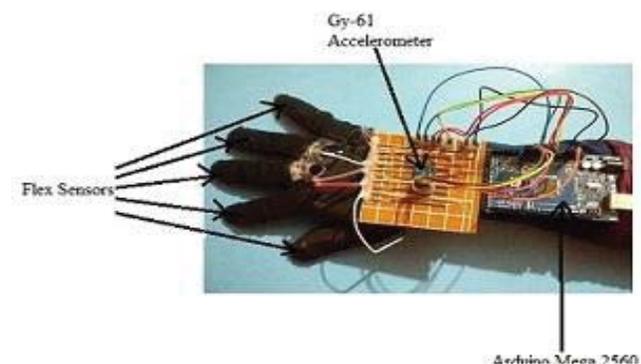


Figure 6: Glove Controller

List of Signs our system is train to predict

Sign	ASL	Sign	ASL
	A		AND
	B		NO
	C		EAT
	D		I NEED TO GO TO WASHROOM
	E		I
	OKAY		STOP
	HELLO		THANK YOU
	CAN		BYE

Some Dataset Values for Some Gestures of our dataset

Index	SVTHUMB	SVINDEX	SVMIDDLE	SVRING	SVPINKEY	X-axis	Y-axis	Z-axis	Class
0	277	68	131	131	188	306	339	354	A
1	281	71	139	131	174	303	321	385	A
2	272	88	142	141	185	302	363	348	A
3	286	78	143	141	179	309	323	338	A
4	270	109	160	158	203	308	350	353	A
5	288	66	148	135	198	304	309	383	A
6	271	75	133	143	199	304	309	386	A
7	246	78	144	135	151	308	333	350	A
8	272	68	139	128	151	308	333	370	A
9	287	84	133	144	172	304	309	388	A
10	251	74	148	134	181	308	333	350	A

Figure: Dataset Values for alphabet A

Index	SVTHUMB	SVINDEX	SVMIDDLE	SVRING	SVPINKEY	X-axis	Y-axis	Z-axis	Class
208	248	165	184	182	214	388	351	367	C
209	224	158	179	175	208	392	351	349	C
210	234	163	182	179	226	393	352	361	C
211	218	143	173	165	218	392	352	362	C
212	226	162	180	176	217	393	352	344	C
213	223	148	176	178	204	392	351	362	C
214	228	156	180	175	218	393	352	367	C
215	196	125	151	136	189	392	356	363	C
216	213	143	163	157	218	381	358	365	C
217	181	135	168	142	199	395	353	357	C

Figure: Dataset Values for alphabet C

Index	SVTHUMB	SVINDEX	SVMIDDLE	SVRING	SVPINKEY	X-axis	Y-axis	Z-axis	Class
816	269	219	242	248	264	398	366	351	Hello
817	268	219	240	240	259	398	342	348	Hello
818	266	221	244	250	284	394	310	345	Hello
819	264	223	242	252	275	396	382	350	Hello
820	262	207	239	228	252	399	322	340	Hello
821	268	217	238	239	249	388	388	343	Hello
822	259	215	239	238	264	397	349	348	Hello
823	253	219	244	242	267	392	384	355	Hello
824	275	219	240	243	257	376	378	352	Hello
825	273	211	241	238	265	393	352	350	Hello

Figure: Dataset Values for Word Hello

Index	SVTHUMB	SVINDEX	SVMIDDLE	SVRING	SVPINKEY	X-axis	Y-axis	Z-axis	Class
586	226	119	255	277	328	395	544	351	Okay
587	229	124	254	269	313	391	347	372	Okay
588	234	132	255	272	318	386	349	355	Okay
589	232	125	257	274	311	391	358	369	Okay
510	238	131	257	265	294	388	345	352	Okay
511	226	132	238	236	282	386	342	368	Okay
512	229	131	249	269	299	388	349	378	Okay
513	224	127	263	278	319	387	346	375	Okay
514	224	103	259	275	311	391	343	371	Okay
515	187	121	233	242	281	390	361	357	Okay

Figure: Dataset Values for Word Okay

VII. Conclusion

Smart gloves which take input from the users' gestures and gives a textual output. It helps the mute people to communicate with one another and with the normal people. The mute people use their standard sign language to communicate which is not easily understandable by common people without the knowledge of the sign language. Thus, this system converts the sign language into audio, pictorial and into text form, improving their life in a significant way. Image processing can be applied to take in new gestures from the user themselves and store it in the dataset. Make proper, well finished and a wireless wearable technology to bridge the communication gap with accurate and precise result.

References

- [1] Hope M Hurlbut, "A Lexicostatistic Survey of the Signed Languages in Nepal," SIL Electronic Survey Reports, p. 23, June 2012.
- [2] X Chai et al., "Sign Language Recognition And Translation with Kinect," Institute of Computing Technology, CAS, Microsoft Research Asia, Beijing, China, October 2013.
- [3] YQuan and PJinye, "Application of improved sign languages recognition and synthesis technology," Industrial Electronics and Application, 2008, ICIEA 2008, 3rd IEEE Conference, pp. 1629-1634, June 2008.
- [4] D Jun, "Gesture Recognition," Shan Hai Maritime University, China, May 2004.
- [5] Ata Ur Rehman, Salman Afghani, Muhammad Akmal, and Raheel Yousaf, "Microcontroller and Sensors Based Gesture Vocalizer," Rachna College of Engg. and tech., Air University Islamabad, Islamabad, Research Paper 2008.
- [6] Prerana K.C, Nishu Mahato, and Bishruti Satyal, "Hand Gesture Recognition using KINECT," Nepal Engineering College, Baktapur, Bachelor Thesis P3-I.0-14-4, 2015.
- [7] M. S Kasar, Anvita Deshmukh, Akshada Gavande, and Priyanka Ghadage, "Smart Speaking Glove-Virtual tongue for Deaf and Dumb," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 5, no. 3, p. 7, Mar. 2016.
- [8] Sparkfun. (2015) learn.sparkfun.com. [Online]. <https://learn.sparkfun.com/tutorials/flex-sensor-hookup-guide/all>.
- [9] Robotpark. robotpark.com. [Online]. <http://www.robotpark.com/GY-61-DXL335-3-Axis-Accelerometer-Module>.
- [10] Arthur Samuel, "Some Studies in Machine Learning Using the Game of Checkers," IBM Journal of Research and Development, vol. 3, no. 3, pp. 210-229, 1959.
- [11] T. K. Ho, "The Random Subspace Method for Constructing Decision Forests," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 8, pp. 832-844, 1998.
- [12] Savan Patel. (2017, May) medium.com. [Online]. <https://medium.com/machine-learning-101/chapter-5-random-forest-classifier-56dc7425c3e1>
- [13] Avinash Navalani. (2018, May) datacamp.com. [Online]. <https://www.datacamp.com/community/tutorials/random-forests-classifier-python>

* * *