DOI: 10.14421/ijid.2021.2384 IJID (International Journal on Informatics for Development), *e-ISSN: 2549-7448*
Vol. 10, No. 2, 2021, Pp. 92-101

A Sign Language Prediction Model using Convolution Neural Network

Rebeccah Ndungi
Informatics, Master Degree
UIN Sunan Kalijaga Yogyakarta
Yogyakarta, Indonesia
rebeccahndungi94*@gmail.com*

Samuel Karuga
Computer Science Department
St Paul's University
Limuru, Kenya
skaruga@spu.ac.ke

Article History
Received November 22nd, 2021
Revised January 15th, 2022
Accepted January 25th, 2022
Published February, 2022

***Abstract*— The barrier between the hearing and the deaf communities in Kenya is a major challenge leading to a major gap in the communication sector where the deaf community is left out leading to inequality. The study used primary and secondary data sources to obtain information about this problem, which included online books, articles, conference materials, research reports, and journals on sign language and hand gesture recognition systems. To tackle the problem, CNN was used. Naturally captured hand gesture images were converted into grayscale and used to train a classification model that is able to identify the English alphabets from A-Z. Then identified letters are used to construct sentences. This will be the first step into breaking the communication barrier and the inequality. A sign language recognition model will assist in bridging the exchange of information between the deaf and hearing people in Kenya. The model was trained and tested on various matrices where we achieved an accuracy score of a 99% value when run on epoch of 10, the log loss metric returning a value of 0 meaning that it predicts the actual hand gesture images. The AUC and ROC curves achieved a 0.99 value which is excellent.**

Keywords*— *Classification; communication gap; deaf communities; hand gestures; Kenyan Sign Language.

1 INTRODUCTION

Imagine a country where we can communicate through video using sign language. A computer can often be programmed to assist individuals in quickly grasping a few sign language gestures that may help them converse with hearing people. Sign language is the most natural method of communication [1], and a computer can often be programmed to assist individuals in quickly grasping a few sign language gestures that may help them converse with hearing people. Linguistic communication, according to [2]. It is a visual-gestural language used by the deaf. To transmit messages, three-dimensional areas and hand movements (along with various body parts) are used. In terms of communication, the SL helps to bridge the gap between the deaf and the public. Ref. [3] claims that non-verbal communication has its own lexicon and syntax distinct from spoken and written languages. In spoken languages, the address schools are used to map sounds to words and grammatical combinations to communicate relevant information. Sign language hand movements use different visual schools than spoken language. To give comprehensive communications, speech communication employs rules; similarly, language communication is governed by a fancy descriptive linguistics.

Due to the communication gap between hearing and deaf individuals, deaf people in Kenya have experienced numerous obstacles in the job and education sectors over the years. Ref. [4] discusses how the educational diversity of Kenyan communities should be evaluated in order to accommodate the deaf. The deaf require a more realistic view of linguistics in their education. Kenya is a multilingual country, with the majority of its residents speaking more than two languages, particularly English, Swahili, and their home tongue, and many people have a propensity to lump all Kenyan Sign language users together. While the majority of Kenyans speak three languages, the deaf group speaks only one. They usually communicate in KSL, and others who do not know SL make it difficult for deaf people to interact effectively. The deaf sign language is regarded as a minority language, and as such, there is a need to conserve and preserve its cultural and linguistic uniqueness, which distinguishes them from the others. Because SL is one of the languages on the edge of extinction, its preservation and identification are critical for history books.

Models and techniques for sign language translation have been developed over time. As seen in Figure. 1, [5] devised a technology that transforms sign language to text automatically. This technology comprises of a glove that a deaf person can wear to communicate with hearing persons in real time. The system converts the signs into letters, which is subsequently displayed on a computer, using Bluetooth connectivity. This is done by converting the signals to digital data and comparing it to a lookup database to get the final letter. The system was designed to enable the deaf converse without a human interpreter.

This research will use a Convolutional neural network technique in conjunction with Tensor-flow [6], a Google Open Source Machine Learning Framework for data-flow programming across a variety of tasks where tensors (arrays) communicate with one another. To apply the Tensor-flow + CNN technique, the classifier must use known samples.

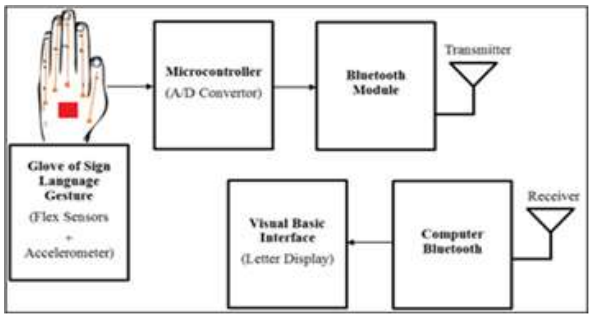

Figure 1. Finger Spelling Sign Language Translator Glove System ([14])

In this situation, known examples from the data set will be used throughout the CNN algorithm's training phase. The prediction should be able to predict the fed-in hand gestures once it's finished.

1.1 Outcomes of this study

Recognizing gesture is an interesting machine vision challenge that is also immensely valuable for deaf folks who want to interact with people who don't understand the Sign Language. Communication is such a vital component of society, the results of this study will considerably aid the society. The rising desire for equality, as well as the lack of a communication gap, justifies the need for more efficient, life-changing communication methods. We live in a world where information is fully networked. The globe has become a global community as a result of rapid technology breakthroughs. At the international level, information sharing among varied groups in society has become extraordinarily easy, effective, and efficient. You can effortlessly access any information according to your needs and preferences by clicking a simple button on your computer. You can't envisage a universe or a condition in which ideas, sentiments, affections, responses, assertions, facts, and numbers aren't exchanged. Communication has been one of the most fundamental parts of human life from the dawn of time. You can effortlessly access any data according to your needs and preferences by clicking a simple button on your computer. Language has become one of the most fundamental parts of life as from ancient period. Strong and effective communication lines have enabled global economic integration. Communication has altered tremendously in recent years. Countries with very well information systems networks today wield economic power. Understanding it as a process, structure, dialogic foundation, and structure involves discussion.

1.2 Problem Statement

People with hearing impairments are less likely to be considered for video consultations, and they are rarely shortlisted for various job sectors; resource distribution is centered on physical appearance rather than equity standards. Ref. [7] illustrates how deaf people feel lonely and alone. With the increasing use of technology, deaf individuals must be able to converse naturally with hearing people and be taken into account in all government areas. They must be able to perform duties in any industry, and deaf children must attend

regular international schools, as opposed to well-performing youngsters who must attend a local special school owing to their physical appearance.

The goal of the research is to create a computerized sign language and recognition model for Deaf people. For a particular hand gesture, the model should automatically translate and show an alphabet. This should ensure that there is no barrier to communication between the deaf and hearing communities. A community where deaf individuals can work in any field and their opinions are valued, and where deaf children can attend regular schools, ensuring that there are no special schools and no restrictions on resources and possibilities owing to physical characteristics.

1.3 Study Objectives

The study goes through a step-by-step process to achieve the goals listed below. Data gathering was employed for the model training, with each American Sign Language letter containing 3000 photographs for training and testing, which was done according to an 80-20 rule. The project was then developed in the second step. After that, the project was put through its paces. Finally, the project was released to the public. However, due to the dynamism of sign language and its dialects, upkeep will be carried out with adjustments.

1.4 Justification

This research will benefit the Kenyan deaf population by bringing equity to the job market, resource allocation, and making them feel accepted by guaranteeing that they are treated similarly. It will also bridge the communication gap between the hearing and deaf cultures. In parent-child relationships and video data collection scenarios, for example. This strategy can be implemented in all schools to teach pupils sign language. This would allow the government to reap significant benefits while also gradually improving and adding to the communication industry.

1.5 Previous Research

Various people have suggested and developed the following projects related to hand gesture recognition and improving human-computer interaction (HCI).

1.5.1 *Hand Gesture Recognition Using Kinect:* Hand surveillance and activity recognition are difficult problems: how to reliably identify the hand and recognize the hand motion. Figure 2 shows the system's framework. Unlike other systems that use color markers to recognize hand shapes, the process utilizes the Kinect sensor's depth map and color information. There is no crowded backgrounds. Timers represent fragmented hand shapes [8]. Gesture detection is still a challenge with the Kinect sensor. Kinect sensors typically have a 640480 resolution. Small objects, such as a human hand, are harder to identify and divide in images of this quality. Thus, the Finger-Earth Mover's Length is used to consider various hand forms [9]. The metric

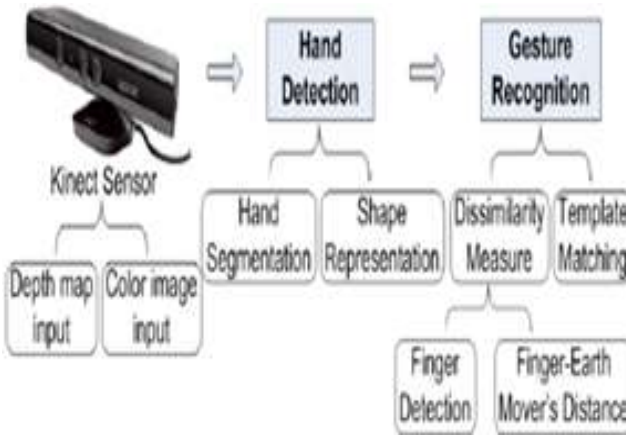

Figure 2. System's framework utilizing Kinect sensor's depth map and color information

is designed to match hand shape differences. Work is required to shift dirt mounds plus the penalty for unmatched fingers equals the heterogeneity distance of two hand shapes. In [9] a near-convex form deconstruction approach from [10] is used. In the end, the system selects a weapon at random. So the system may choose between computers and humans. In this demo, a Kinect sensor was used to recognize hand gestures efficiently and accurately. Hand detection uses both depth and color information from the Kinect sensor, ensuring robustness in cluttered environments. Moreover, "gesture recognition module's Finger-Earth Mover's" Distance metric effectively recognizes hand shapes with input variations and distortions. This hand gesture recognition system works well in two real-world HCI applications and can be used in many other hand gesture-based HCIs.

1.5.2 *Quantitative Doppler Characterization for High Precision Gesture Sensing:* The proposed method makes use of the Doppler Effect [11]. The technology collects a "pilot tone from the device's speaker (origin) (receiver)". The body reflects waves as it approaches the device, it causes a frequency change. This is because the transmitter and destination are both fixed. When you leave the device, the frequency goes negative. The receiver calculates shifts using the expression shown in Formula 1:

$$f = \left(1 + \frac{\Delta v}{c}\right) f_0 \quad (1)$$

It depicts perceived frequency at the microphone and speaker, sound speed in air, and acceleration of an in-air pose roughly comparable to the gadget. The hand's wavelength shifts are plotted. An intensity is emitted when no action is undertaken. Doppler phenomenon occurs when a hand advances or flees from a sensor.

Quantification of the Doppler Effect: This phase analysis it shows the Doppler shift from the raw voice input. A speaker emits a tone, and a microphone captures it, therefore this stage starts with signal collection and finishes with feature characteristic quantization. The sub-modules are listed below.

Data Gathering

For the system, a pure sinusoidal wave played through the device's sound is sufficient. Most laptop and phone speakers can emit audio up to 22 kHz. Inaudible to persons yet traceable by practically all common electronics, an 18 kHz pitch was employed. The NY Quist theorem required that we record the mike input at 44.1 kHz. Every data frame took 50 milliseconds to acquire [12] Jerking motions in front of a pc were captured at speeds up to 3.9 m/sec [13].As a result, every gesture yields five pieces of raw data. The frequency-shift occurring in the frequency domain is the main effect of human hand actions. As a result, the original time domain signal has to be converted to frequency domain using the Fast Fourier Transform (FFT)

Gesture Recognition

In this module, segmentation was conducted. It is an acknowledging when a user makes a motion is vital to system functionality. It conserves energy by reducing accidental actions/noises. We use the parameters of feature primitives to see if the recurrence fluctuates in a frame time. The gesture starts when the primitive is not zero. Four consecutive zeroes indicate that movements have ended and a gesture will be returned instantly. This method is appropriate for real-time extraction of motions. If no motion is detected for a long time, the system turns off automatically.

The HMM is a stochastic data analysis model [13]. It is used in voice recognition and gesture recognition systems [14]. Feature semantics in this system are vectors 0. The feature vector V is used to train the classifier and recognize gestures. The method runs through various numbers of disguised states until it finds the most accurate. For each move, the system builds an HMM. Each HMM is specified by a triplet (B Q V).

Where B is a Stage Switchover, matrix Q is a Sentiment matrix after instantiating the classifiers for each motion, the training technique employs V as an observation vector. The Baum-Welch algorithm refines constants until models are ideal. Using the trained HMMs, we evaluate the probability of each sighting and produce the gesture with the highest score.3. This is the initial step in mapping app actions to library gestures. Gestures can be used as a game pad. A right swipe moves a character to the right, whereas a tap jumps a character. Users can also create and perform their own custom gestures. The system also provides the gesture's properties to the app. For example, velocity fluctuates over time and can be used to control a character's movement. It offers an ultrasonic Doppler effect-based hand gesture recognition system. It can detect a wide stretching without modifying a laptop. One of the system difficulties we resolved was similar motion detection.

The suggested technique detects similar gestures in a greatly larger gesture library with 95% accuracy. Using six basic kinematics and HMM categorization, our technique improves accuracy to 98.6%. These findings highlight our system's resilience and adaptability to different environmental conditions, paving the way for future non-contact gesture-based human-computer interaction. A single model can be used by different users, but the correctness will be lowered.

1.5.3 Vision-based Portuguese Gesture Recognition ClassificationI: In order to recognize the Portugal Gesture Recognition alphabet in Figure 3, the prototype uses real-time vision. The prototype tested and selected hand traits that may be used with machine learning techniques for real-time sign language recognition. This involves making sign language movements in front of the camera, which the system reads and categorizes on the interface's right side. The implemented approach uses only one sensor, a Kinect webcam [15], and is dependent on several hypotheses.

The recommended system design contains two modules: data capture, pre-processing, and extraction of features; and sign language gesture categorization. The first module recognizes, tracks, and segments the hand in video. The split hand provides features for categorizing gestures. After that, the gesture categorization method includes a previously trained SVMs (SVM), a pattern recognition technique in supervised classification that performs well with high-dimensional facts.

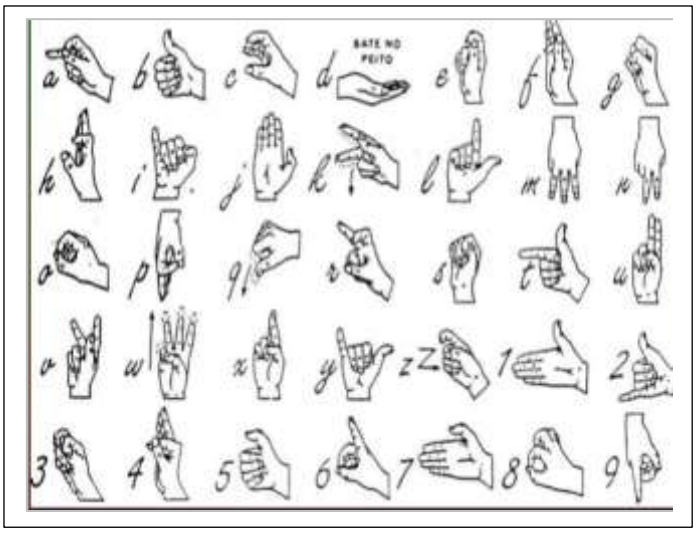

Figure 3. The Portugal Gesture Recognition alphabet

The prototype's HCI was designed in C++ using the openFrameworks toolkit [16] with the OpenCV [17] and OpenNI [18] addons. OpenCV handled several vision-based tasks including identifying the hand blobs structure, whereas OpenNI handled RGB and depth picture acquisition. The free source Dlib library was used for training the model and gesture categorization [19]. The final implementation of the SVM classifier was chosen because to its excellent accuracy in the trials utilizing the chosen variables. The resulting model was also compact and rapid, making it suitable for real-time classification applications.

This experiment compared two types of hand features to discover which one performed better in Portugal Gesture Recognition. These characteristics are derived from the body perimeter dimensions and are also referred as shape signatures. As per [20] and [21], it performs well in form retrieving and categorization while being computationally straightforward.

This study used a SVM Classifier to examine two types of hand attributes for gesture recognition (SVM). This learning strategy was chosen because it has previously produced good classification results using first-hand features. Table 2 reveals that the categorization accuracy of the two possible parameters was only 0.2 percent different. Based on the observations so far in this, it can be found that the present characteristics are not beneficial, as they increase the number of properties and the final data set and model sizes. With regarding to categorize the motions, the confusion matrix shows high categorizing errors between gesture one and gestures three and five (corresponding to the 'E' and 'U' vowels, respectively). However, the centroid distance is simple to compute, resulting in lower data sets and final model files.

The results are very promising, but further testing is needed to assess the suggested system's effectiveness and the convolutional model's accuracy (Table 1).

Table 1 Table Type Styles

Parameters	data set 1	data set 2
Kernel type	Linear	Linear
C	1	2
Accuracy	99.4%	99.6%

2 METHOD

In the computer vision industry, developing accurate Machine Learning Models capable of detecting and localizing many objects in a single image has remained a major challenge. CNN is one of the most basic machine learning techniques [22]. In many machine - learning issues, the Convolutional Neural Network (CNN) has showed outstanding performance. Many people in the industry love it because of how simple it is to use and how fast it calculates. Many real-world applications of algorithms can be found in recommendation systems and classification tasks based on similarity. This approach categorizes data points depending on how similar they are to each other. It uses test data to generate an "informed judgment" on how an unlabeled point should be classed as the pattern of connectivity in a CNN stems from their research on the visual cortex's architecture. A convolutional neural network has many layers, including convolution, pooling, and fully connected layers, and uses back propagation to learn spatial hierarchies of features.

Pre-processing data is likely to be noisy, inconsistent, and incomplete, or it contains missing data. Such data can have a significant impact on the accuracy of data mining results, necessitating pre-processing, or rather, preparing and modifying the data in order to obtain highly accurate results [6]. As a result, data pre-processing is an important stage in data mining. There are four types of data processing methods: training, data cleaning, data integration, data transformation, and data reduction.

2.1 Data Training

The data set utilized, which is freely available on kaggle.com, is modeled after the famous Modified National Institute of Standards and Technology data set (MNIST). Each training data set divided into alphabets A-Z, with each alphabet including 3000 raw photographs marked A1, A2, A3, ..., Z3000, while the test data set also included the 26 alphabets (pictures) in raw form named A test, B test, C test, and so forth.

Data split between training and testing in order to correctly execute the ML method. This is because a machine-learning algorithm operates in two stages when working with data sets: testing and training. The data manually separated into test-train 20 percent - 80 percent in this project, with a training data set including 80 percent training data and 20% testing data. The data set was imported using pandas, and the training function was performed using the Tensor Flow library. To enable trans-version over directories, the OS library is utilized. The RANDOM library allows for image shuffling during training. Matplotlib [23] is crucial since it aids in the visualization of data via graphs. The classes are stored in an array CATEGORY [ ], with the content being "A," "B," and "C." The X and Y arrays are appended with the label and features. By dividing X by the maximum number of pixels, 255, the data is normalized. The model is built using three convolution layers, each with a batch size of 32, epochs of 10, and a validation split of 0.1. The dropout is used to avoid over-fitting and is set at 0.3. Notice that the train data comprises 9000 points as seen in code snippet appended below, while the test data has 2400 points, or 20% of the original data, to deliver the exact result needed to complete the project. Check out the snippet below in listing for a short look at the training function as well as the cell result on train data, which is 9000.

```
DATADIR="C:\\Users\\\\Desktop\\Project
  ↪data set\\data set\\asl_alphabet_train\\asl_alphabet_train"
```

```
    CATEGORIES=["A","B","C"]
    # The size of the images that
    your neural network will use
    IMG_SIZE=50
    # Checking or all images in the
    data folder
forcategoryinCATEGORIES:
    path=os.path.join(DATADIR,
    category)
    forimginos.listdir(path):
            img_array=cv2.imread(os.path
    .join(path,  img),
    cv2.IMREAD_GRAYSCALE) training_data=[]
defcreate_training_data():
    forcategoryinCATEGORIES :
      path=os.path.join(DATADIR,
    category)
    class_num=CATEGORIES.index(category)
    forimginos.listdir(path):
      try:
      img_array=cv2.imread(os.path.jo
    in(path,  img),  cv2.
      ↪IMREAD_GRAYSCALE)
      new_array=cv2.resize(img_array,
    (IMG_SIZE,  IMG_SIZE))
    training_data.append([new_array,
    class_num])
exceptExceptionase: pass

    create_training_data()

print(len(training_data))

    9000
```

2.2 *Cleaning up the data*

Data have missing values, it indicate that the data are incomplete; this may occur when data were collected in a rush or when some respondents declined to provide certain details; data were noisy, it indicate that the data contain errors or outliers, causing confusion; and data contain duplicates. Data cleaning removes outliers, handles missing numbers, and deals with data that is inconsistent. We've fine-tuned the data set we're utilizing, and there are no missing photographs.

2.3 *Integration of data*

Data integration is the process of combining data from several sources and storing it in a single location. This can be a hurdle because we are unsure of some things, therefore data integration was not done in the data-set because it was comprehensive and contained all of the properties, so there was no need.

2.4 *Data Transformation*

Data transformation involves preparation of the data set in order to produce a data set used for training of the model. The pictures used in the model training, were in their natural form and needed to be transformed into gray-scale. The data transformation into gray-scale enables faster training and make the imaged to be uniform.

2.5 *Data reduction*

Large-scale data mining is inaccurate and time-consuming. Data reduction is useful since it reduces data set size while retaining useful information. The smaller data set allows for better mining. In this situation, the photos were trimmed to lower the training data volume.

2.6 *Testing*

The test-set used to compare models, when employed in the actual world; it contains carefully sampled data from several classes. To show the alphabet location, the model predicts an alphabet in a category and prints out a true. In Figure 4, the model was tested using the letter A and it successfully printed a true when a match was discovered.

Figure 4. Model testing using the letter A

3 RESULT AND DISCUSSION

Hearing impaired people face a variety of challenges, which can be divided into three categories: fewer educational and employment opportunities due to impaired interaction, social phobia due to limited access to public services difficulties communicating with others, and emotional problems due to a loss of self-esteem and confidence caused by a communication barrier between them and hearing people. With the increasing usage of technology, deaf people must be compelled to be able to communicate naturally with hearing people and be considered as a whole by government sectors. The goal of this project is to use sign language to take a first step in breaking down communication barriers between normal people and deaf and dumb persons.

The main objective was to create a model able to predict a sign language hand gesture captured and translate it into English alphabets. The CNN machine-learning algorithm used together with the tensor flow was tested on three different metrics. Matrices are an important part of linear

algebra. Data from Kaggle, a public open internet data set library for ML tools, was used in the proposed research. The file format is modeled around the classic NIST format MNIST. Each trained data set is divided into alphabets A-Z, with each alphabet having 3000 raw photos marked A1, A2, A3, ..., Z3000, a space hand gesture marked A1, A2, A3….. A3000 and a delete too marked similar. Whereas the test data set includes the 26 alphabets (pictures) in raw form labelled A test, B test, C test... Z test as seen below on Figure 5, 6 and 7 respectively.

Figure 5. A1 raw hand gesture

Figure 6. S1 space hand gesture

Figure 7. Delete1 raw hand gesture

The pictures were then converted into grayscale so as to make them uniform through dimension reduction, reduction on algorithm complexity besides that, some algorithms only work best with grayscale images. In the discipline of machine learning, matrices are used to define procedures and processes, such the training data variable (X). The algorithm was tested for credibility based on the following metrics; accuracy score, logarithmic loss and AUC. The AUC is highly preferred because the AUC is unaffected by scaling. Rather of measuring real numbers, it assesses how well forecasts are ordered and because the UC is not affected classification criterion it analyzes the models forecast accuracy independent of categorization level.

Under the AUC, algorithm was first tested with an epoch 5, where the prediction was 70.94 while on a 10 epoch, the prediction was 99.94. The log loss metric, returned a log loss value of 0.017069. A log loss of below 0 is considered a strong algorithm while on AUC, the value was 1.The ROC curve gave a value of 0.99 value which is considered excellent while the confusion matrix was able to predict and visualize the predicted classes against the actual values.

These algorithm can be used by various researchers in the fight to fill the communication gap between the hearing and the deaf communities in the country. The model was subjected to more than three evaluation matrices and all turned out with excellent and acceptable values and results making the model great for use and implementation. Such a system can be applied in the different sectors in the country such as schools where both the hearing and the deaf communities can attend with much ease.

3.1 Accuracy Score

The accuracy of classification models is one parameter. Informally, accuracy refers to our model's accuracy rate. The formal definition of accuracy is given in Formula 2:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (2)$$

Classification accuracy was the first metrics and it is the number of correct predictions divided by the total number of predictions multiplied by 100. A case of a convolution neural network prediction accuracy score: 0.9994 * 100 =99.94 percent. This result was after changing the epoch value to 10 from 5 where an iteration of 5 gave 70.94. Figure 8 shows a quicker insight with an epoch = 10 iteration.

```
Epoch 1/10
162/162 [==============================] - 14s 81ms/step - loss: 0.9965 -
accuracy: 0.4351 - val_loss: 0.5238 - val_accuracy: 0.6711
Epoch 2/10
162/162 [==============================] - 13s 80ms/step - loss: 0.5145 -
accuracy: 0.6628 - val_loss: 0.4604 - val_accuracy: 0.6756
Epoch 3/10
162/162 [==============================] - 13s 79ms/step - loss: 0.4456 -
accuracy: 0.7051 - val_loss: 0.3728 - val_accuracy: 0.7744
Epoch 4/10
162/162 [==============================] - 13s 80ms/step - loss: 0.3545 -
accuracy: 0.8363 - val_loss: 0.2815 - val_accuracy: 0.9589
Epoch 5/10
162/162 [==============================] - 14s 85ms/step - loss: 0.2650 -
accuracy: 0.9420 - val_loss: 0.2511 - val_accuracy: 0.9478
Epoch 6/10
162/162 [==============================] - 15s 94ms/step - loss: 0.2091 -
accuracy: 0.9599 - val_loss: 0.1550 - val_accuracy: 0.9789
Epoch 7/10
162/162 [==============================] - 13s 83ms/step - loss: 0.1585 -
accuracy: 0.9798 - val_loss: 0.1360 - val_accuracy: 0.9800
Epoch 8/10
162/162 [==============================] - 13s 79ms/step - loss: 0.1287 -
accuracy: 0.9838 - val_loss: 0.0999 - val_accuracy: 0.9889
Epoch 9/10
162/162 [==============================] - 13s 80ms/step - loss: 0.0999 -
accuracy: 0.9916 - val_loss: 0.1030 - val_accuracy: 0.9778
Epoch 10/10
162/162 [==============================] - 13s 81ms/step - loss: 0.0901 -
accuracy: 0.9875 - val_loss: 0.0764 - val_accuracy: 0.9911
Saved model to disk
```

Figure 8. Epoch = 10 iteration

3.2 Logarithmic Loss

Given by the Formula 3:

$$Logloss = \frac{1}{N}\sum_{i=1}^{N} logloss_i \quad (3)$$

This metric assess the classification model's performance. It determines how far the projected probability differs from the actual label [23]. As a result, the lower the log loss number, the more flawless the model. Log loss value = 0 for a perfect model. Figure 9 shows a log loss of 0.017069407234533145 obtained. Log loss value = 0

```
#model log loss
fromsklearn.metricsimportlog_loss
loss=log_loss(y_test, y_pred, eps=1e-15,
normalize=True, sample_weight=None,
←labels=None)
print(loss)
0.07638954399838868
```

Figure 9. Log loss value of 0

3.3 AUC (Area under the Curve)

The ROC curve is a binary classification issue evaluation metric. You can use it to separate signal from noise using a probability curve that compares TPR to FPR as shown in Formula 4 and Formula 5 respectively at various threshold levels. One can see TPR v/f at various classification levels. Lowering the classification criterion increases both False and True Positives.

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

$$FPR = \frac{FP}{FP+TN} \quad (5)$$

AUC assesses performance across all levels. AUC is the likelihood that the model ranks a positive example higher than a negative one. AUC is a 0-1 number. AUC of 0.0 for 100% erroneous predictions and 1 for 100% right predictions. As seen in Figure 7 where 1 was achieved while figure 8 shows the ROC of class B which is commonly used to evaluate Boolean classification techniques. Unlike most other measures, it presents a graphical picture of a classifier's success with a 0.99 value which is considered excellent (Figure 10, and Figure 11).

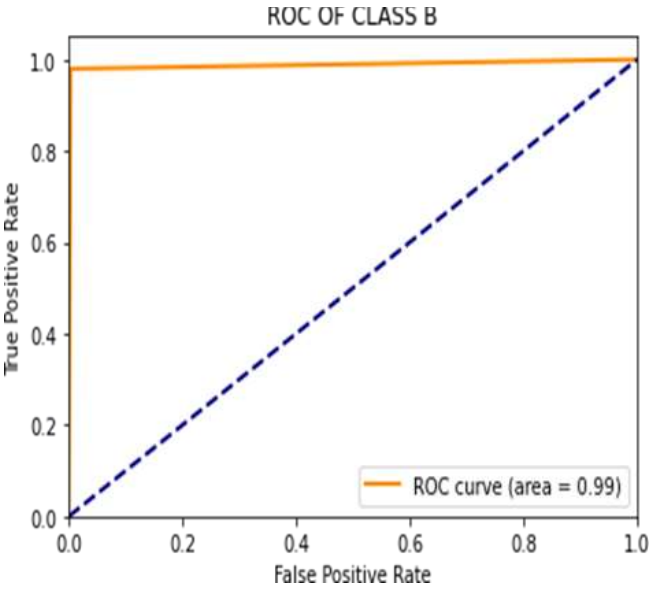

Figure 10. Area under Curve (AUC)

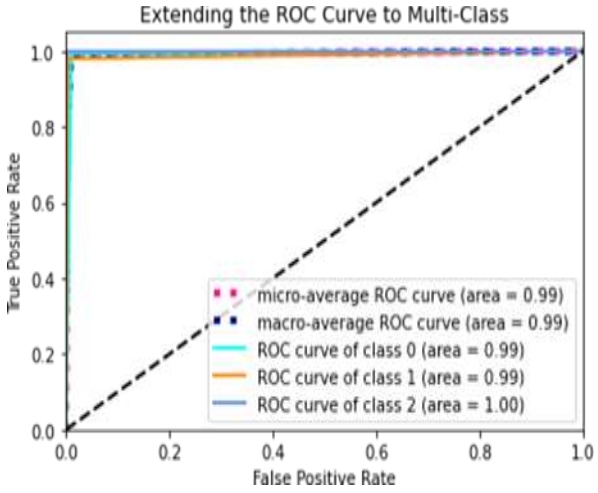

Figure 11. Extended ROC curve to multi-class.

3.4 Confusion Matrix

This is an N*N matrix used for performance evaluation on classification models with N being the number of target class. The matrix performs a comparison of the ML predicted values against the real values. Rows are the predicted value. In Figure 12, the model was evaluated on the confusion matrix and produced the below heat map results which are true positive.

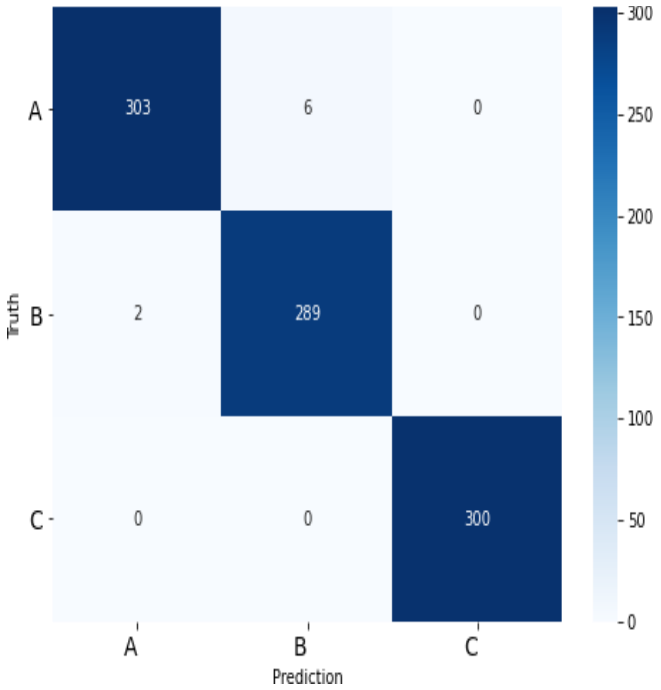

Figure 12. Confusion matrix heat-map

3.5 Model "Accuracy" and "Loss"

Accuracy is a performance metric for classification models and it is normally in percentage (%). Accuracy is the number of predictions that match the true value as seen on Figure 13 while the loss is a prediction measured by how much it differs from the true value. It is the sum of all errors made in trained or validated sets as seen on Figure 14.

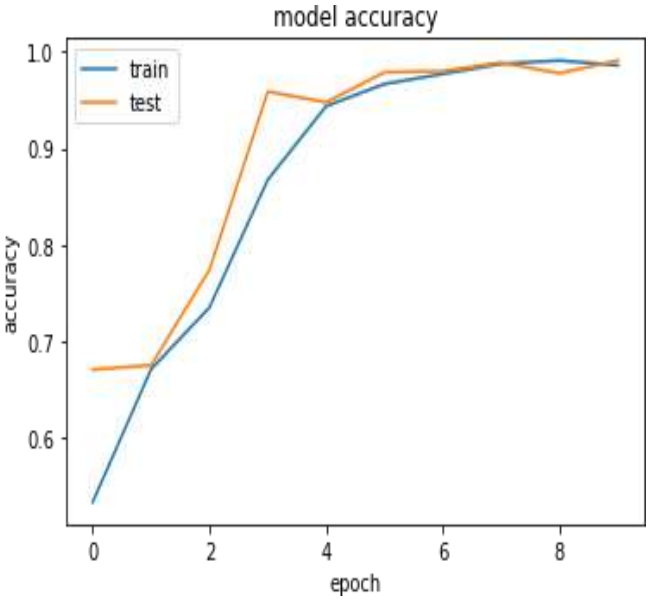

Figure 13. The model accuracy

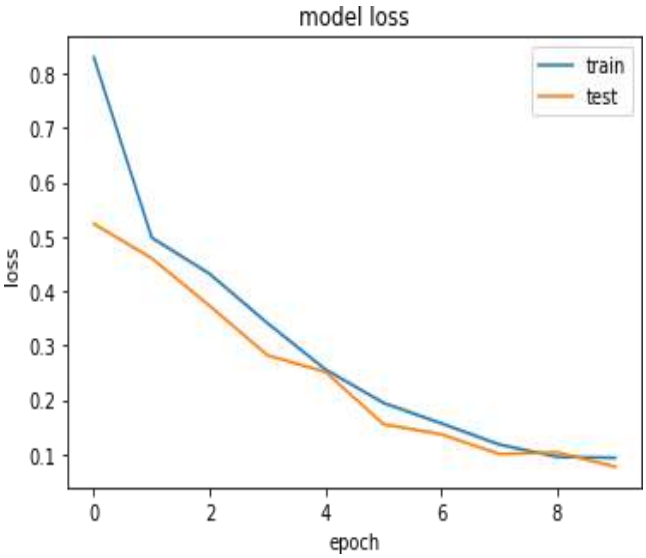

Figure 14. The model loss

3.6 Prediction

Prediction is very difficult especially about the future. There was the need to design a predictive model for sign language prediction with more improvements in accuracy using various data mining techniques on a data set collected. Performance of the models was evaluated using metrics of accuracy score, logarithmic loss and AUC metrics in order to achieve results that are more accurate and have less variance. The researcher took more time on data processing especially data transformation and dealing with broken images and non-uniform image sizes, which is an important step in achieving

accurate results. Figure 4 shows that the models performed well on predicting different alphabets.

4 CONCLUSION

This work has revealed the most sign language to English language prediction utilizing data mining approaches. Other researchers who want to develop more consistent research methodologies can use this study's findings. This study uses convolution neural network technique supplied by jupyter notebook in the anaconda navigation platform. The model was tested on various ML evaluation metrics with the accuracy score returning a 99% value when run on epoch of 10, the log loss metric returning a value of 0 meaning that it predicts the actual hand gesture images. The AUC and ROC curves achieved 0.99 value which is excellent and the confusion matrix was able to produce a successful heat-map for the classifier. The researcher plans to conduct more study using more data sets and algorithms to construct models that can anticipate words in English. Further research/work is required in the future.

AUTHOR'S CONTRIBUTION

Sole responsibility for study conceptualization, design, data collection, analysis and interpretation.

COMPETING INTERESTS

This paper is my original work. I did not receive any funding and thus does not have any conflict of interest and the processing of this paper does not relate with my position as a proofreader in IJID.

ACKNOWLEDGMENT

This endeavor would not have been possible without the support of Samuel Karuga, who provided undying support during the duration of completing it. I would also like to thank my adviser/TA for their guidance and support.

REFERENCES

[1] J. G. Mweri, "Diversity in education: Kenyan sign language as a medium of instruction in schools for the deaf in Kenya," *Multilingual Education*, vol. 4, no. 1, p. 14, Aug. 2014.

[2] R. G. Brill and Conference of Educational Administrators Serving the Deaf, *The Conference of Educational Administrators Serving the Deaf: a history*. Washington, D.C.: Gallaudet College Press, 1986.

[3] D. M. Perlmutter, "The Language of the Deaf." Accessed: Nov. 28, 2021. [Online]. Available: https://www.nybooks.com/articles/1991/03/28/the-language-of-the-deaf/

[4] DR. M. G. JEFWA, Mweri, J.G. In Okombo et al. "*Introduction to Theory and Skills of Teaching Kenyan sign Language: A handbook for Teachers Nairobi*," 2006.

[5] X. Teng, B. Wu, W. Yu, and C. Liu," A hand gesture recognition system based on local linear embedding," *Journal of Visual Languages and Computing*, vol. 16, no. 5, pp. 442-454, April 2005.

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

[7] J. Ngugi, S. Kimotho, and S. Muturi, "SOCIAL MEDIA USE BY THE DEAF IN BUSINESS AT NAIROBI, KENYA," *African Journal Of Business And Management*, vol. 4, no. 3, pp. 1-13, 2018.

[8] E. Keogh, L. Wei, X. Xi, S.-H. Lee, and M. Vlachos, "LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures," *VLDB '06: Proceedings of the 32nd International Conference on Very Large Data Bases*, pp. 882–893, Sep. 2006.

[9] Z. Ren, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," *MM '11: Proceedings of the 19th ACM Internasional Conference on Multimedia*, pp. 1093–1096, Nov. 2011.

[10] Z. Ren, J. Yuan, C. Li, and W. Liu, "Minimum near-convex decomposition for robust shape representation," *in 2011 International Conference on Computer Vision*, pp. 303–310, Nov. 2011.

[11] C. Doppler and F. J. Studnica, "Ueber das farbige licht der doppelsterne und einiger anderer gestirne des himmels.," *undefined*, Accessed: Nov. 28, 2021. [Online]. Available: https://www.semanticscholar.org/paper/Ueber-das-farbige-licht-der-doppelsterne-und-des-Doppler-Studnica/62bfc7c7b751aacd02d3b4405ace10aedb2d7987

[12] S. Gupta, D. Morris, S. Patel, and D. Tan, "SoundWave: using the doppler effect to sense gestures," *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, May 2012.

[13] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[14] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, Jan. 1986.

[15] J. R. Chowdhury, "Kinect Sensor for Xbox Gaming," Accessed: 2022. [Online] Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.2368&rep=rep1&type=pdf.

[16] "About open Frameworks." https://openframeworks.cc/about/ (accessed Nov. 28, 2021).

[17] G. R. Bradski and A. Kaehler, *Learning OpenCV: computer vision with the OpenCV library*, 1. ed., [Nachdr.]. Beijing: O'Reilly, 2011.

[18] OpenNI, "OpenNI recognized by PCL as the standard for 3D sensing acquisition." https://www.prnewswire.com/news-releases/openni-recognized-by-pcl-as-the-standard-for-3d-sensing-acquisition-190002261.html (accessed Nov. 28, 2021).

[19] D. King, "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, Jul. 2009.

[20] D. Zhang and G. Lu, "A comparative study on shape retrieval using Fourier descriptors with Different shape signatures," *J Vis Commun Image Represent*, vol. 1, Jan. 2001

[21] P. Trigueiros, F. Ribeiro, and L. Reis, "A Comparative Study of different image features for hand gesture machine learning," *Proceedings of the 5th International Conference on Agents and Artificial Intelligence*, vol. 2, 2013.

[22] "Terms of Use | Kaggle." https://www.kaggle.com/terms (accessed Nov. 28, 2021).

[23] A. Dangi, "'A new approach for user identification in web usage mining preprocessing,'" *IOSR Journal of Computer Engineering*, vol. 11, pp. 57–61, Jan. 2013.