iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52994

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Sign Language Detection Using Deep Learning

Adarsh Srivastava[1], Adarsh Pratap Singh[2], Aditya Anand[3], Ankit Kumar[4], Prof. Prem Prakash[5]
[1, 2, 3, 4]*B.Tech Students,* [5]*Mentor, Dept. of Computer Science and Engineering, Greater Noida Institute of Technology*

Abstract: ***People with hearing impairments are found worldwide; therefore, the development of effective local level sign language recognition (SLR) tools is essential. We conducted a comprehensive review of automated sign language recognition based on machine/deep learning methods and techniques published between 2014 and 2021 and concluded that the current methods require conceptual classification to interpret all available data correctly. Thus, we turned our attention to elements that are common to almost all sign language recognition methodologies. This paper discusses their relative strengths and weaknesses, and we propose a general framework for researchers. This study also indicates that input modalities bear great significance in this field; it appears that recognition based on a combination of data sources, including vision-based and sensor-based channels, is superior to a unimodal analysis. In addition, recent advances have allowed researchers to move from simple recognition of sign language characters and words towards the capacity to translate continuous sign language communication with minimal delay. Many of the presented models are relatively effective for a range of tasks, but none currently possess the necessary generalization potential for commercial deployment. However, the pace of research is encouraging, and further progress is expected if specific difficulties are resolved.***

I. INTRODUCTION

For millions of people, sign language communication is the primary means of interacting with the world, and it is not difficult to imagine the potential applications involving effective sign language recognition (SLR) tools, For example, we could translate broadcasts that include sign language, create devices that react to sign language commands, or even design advanced systems to assist impaired people in conducting routine jobs. In particular, deep neural networks (DNNs) have emerged as a potentially ground breaking asset for researchers, and the full impact of their application to the problem of SLR will likely be felt in the near future. Because hardware and software components have evolved to the point where developing advanced systems with real-time translation capacities appear to be within reach, a large number of exciting and innovative solutions have been proposed and tested in recent years with the objective of building fully functional systems that can understand sign language and respond to commands given in this format.

However, before any truly practical applications can be considered, it is imperative to perfect the interpretation algorithms to the point where false positives are rare.

Owing to the numerous challenges inherent in this task, at this stage, it is not yet possible to design SLR tools that approach 100% accuracy on a large vocabulary. Thus, it is very important to continue developing new methods and evaluate their relative merits, gradually arriving at increasingly reliable solutions. While most researchers agree that deep learning models are the most suitable approach, the optimal network architecture remains a point of contention, with several competing designs achieving promising results. Detailed experimental evaluations are the only way to identify the best performing algorithms and refine these further using discoveries from other research teams when applicable. As most countries use their own variations of sign language, much of the research is conducted locally with persons skilled in using regional signs. It is not surprising that a large number of scientific papers are targeting SLR problems and that the performance level of the proposed solutions is rapidly increasing from year to year.

II. MOTIVATION

"Curiosity about life in all of its aspects, I think, is still the secret of great creative people"

Our project "Sign Language Detection" provides a platform were

1) The Sign Language is very important for people who have hearing and speaking efficiency generally called Deaf and Mute. It is the only mode of communication for such people to convey their messages and it becomes very important for people to understand their language.
2) The motivation behind this work is the possibility of reducing the communication barrier between the deaf and hearing communities.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

III. BACKGROUND

In recent years, there have been ongoing efforts to develop automated methods for the completion of numerous linguistic tasks using advanced algorithms that can 'learn' based on past experience [33]. Sign language recognition (SLR) is an area where automation can provide tangible benefits and improve the quality of life for a significant number of people who rely on sign language to communicate on a daily basis [34]. The successful introduction of such capabilities would allow for the creation of a wide array of specialized services, but it is paramount that automated SLR tools are sufficiently accurate to avoid creating confusing or dysfunctional responses. In this section, we provide a brief background regarding some important approaches that have been utilized for automated SLR.

A. Machine Learning (ML)

The machine learning concept encompasses a number of stochastic procedures that can be used to predict the value of a certain parameter based on similar examples that the algorithm was previously exposed to. A simple example, shows how a general formalization of the learning process takes place. There are many different methodologies that belong to this group; some of the best-known methods include naïve Bayes, random forest, K-nearest neighbour, logistic regression, and the support vector machine. All of these methods undergo a training phase, which can be either supervised (using labelled input data) or unsupervised (without l data), and use input features to establish connections among variables and acquire predictive power. However, owing to their simplicity, such methods have limitations when there is a need to capture nuanced semantic hints, as is the case with most linguistic tasks. On the other hand, they can often provide the foundation for the development of more powerful analytic tools and serve as a measuring stick to evaluate progress. on data input from wearable sensors, which provide a very direct translation of a user's movements. The data can be filtered using techniques such as SVM to provide a reasonably accurate recognition of the intended sign. Some of the aforementioned machine learning methods are used primarily to analyse static content (i.e., individual signs isolated in time and space), while in some cases, there have been attempts to interpret continuous segments of sign language speech, necessitating the use of dynamic models such as dynamic time warping or relevance vector machines. In general, basic stochastic models are better suited for simple SLR tasks, which is why they were extensively used in the early stages of research.

B. Deep Learning

Recently, basic machine learning approaches have been largely replaced with deeper architectures that employ several layers and pass information in vector format between layers, gradually refining the estimation until positive recognition is achieved. Such algorithms are usually described as ''deep learning'' systems or deep neural networks, and they operate on principles similar to the machine learning strategies described above, although with far greater complexity. Based on the structure of the network, two architectures are commonly used for a number of different tasks: recurrent neural networks (RNNs) that include at least one recurrent layer, and convolutional neural networks (CNNs) that include at least one convolutional layer. Depending on the number and type of layers, these networks can exhibit different properties and are generally suitable for different types of tasks, while the training phase decisively impacts the performance of the algorithm. The general rule is that larger and more specific datasets allow for more robust network training, and therefore the quality of the training set is an important factor. Additional fine-tuning of a model can usually be achieved by changing some of the relevant hyper-parameters that define the training procedure . The majority of research involving the automation of SLR tasks is currently based on methods that rely on a combination of images and depth data, which generate a tremendous amount of information that often requires analysis in real time (or at least taking the temporal dimension into account). With larger and more diverse datasets, simple machine learning methods tend to underperform, which is why many of the more sophisticated models are based either on RNN or CNN design. Deep networks can be trained using multimodal input (e.g., skeletal data combined)

IV. METHODOLOGY

This section discusses the proposed methodology for the evaluation of deep models, optimizers and hyperparameters. Before the performance evaluation, the depth data is first processed through depth thresholding to get the hands from the data. Secondly, the segmented hand's binary image is converted to a coloured image for deep models, and lastly, data is augmented to avoid over-fitting. Four pre-trained models viz. InceptionV3, ResNet152V2, InceptionResNetV2 and ReXNet101 are selected from the literature based on their performance in the ImageNet [28] challenge. A customized three-layered CNN model is also designed, trained from scratch, and then compared with the pre-trained deep models mentioned above.

©IJRASET: All Rights are Reserved

This paper uses five gradient-based optimizers: Stochastic gradient descent (SGD), adaptive gradient (AdaGrad), adaptive delta (AdaDelta), root mean square propagation (RMSProp) and adaptive moment estimation (Adam) with their hyperparameters like learning rate, batch size and momentum. This work aims to evaluate the latest deep models, optimizers, and hyperparameters on Indian signs. Thus, two subsets, Numerals and Alphabets, of a publicly available dataset are selected for the comprehensive analysis. First, Numerals subset with 9 classes is used to tune the hyperparameters, which are then applied to the Alphabets subset (24 classes) to evaluate the performance of deep models on a subset using the hyperparameters tuned on a different subset of the same dataset. The process of evaluating the models, optimizers and hyperparameters follows the steps given below.

1) Step 1: This step uses the relatively smaller sized architectureInceptionV3 to tune the batch size on the Numerals subset of the dataset by fixing other hyperparameters and optimizers.

2) Step 2: InceptionV3 is again used to tune the learning rate and momentum (SGD) for each optimizer by fixing the batch size with the value obtained in Step 1 and using the Numerals subset.

3) Step 3: This step evaluates the optimizers with the hyperparameter settings selected from Steps 1 and 2. The optimizer with the lowest loss and highest recognition accuracy on the Numerals subset is selected to evaluate the deep models further. Step 4: Finally, this last step evaluate the deep models with the selected hyperparameters and optimizer from Steps 1, 2 and 3 on both the Numerals and Alphabets of the dataset. Thus, in the end, the CNN model that gives the highest recognition performance on both the subsets of the ISL dataset is considered to be the most suitable for the application of static ISL recognition.

REFERENCES

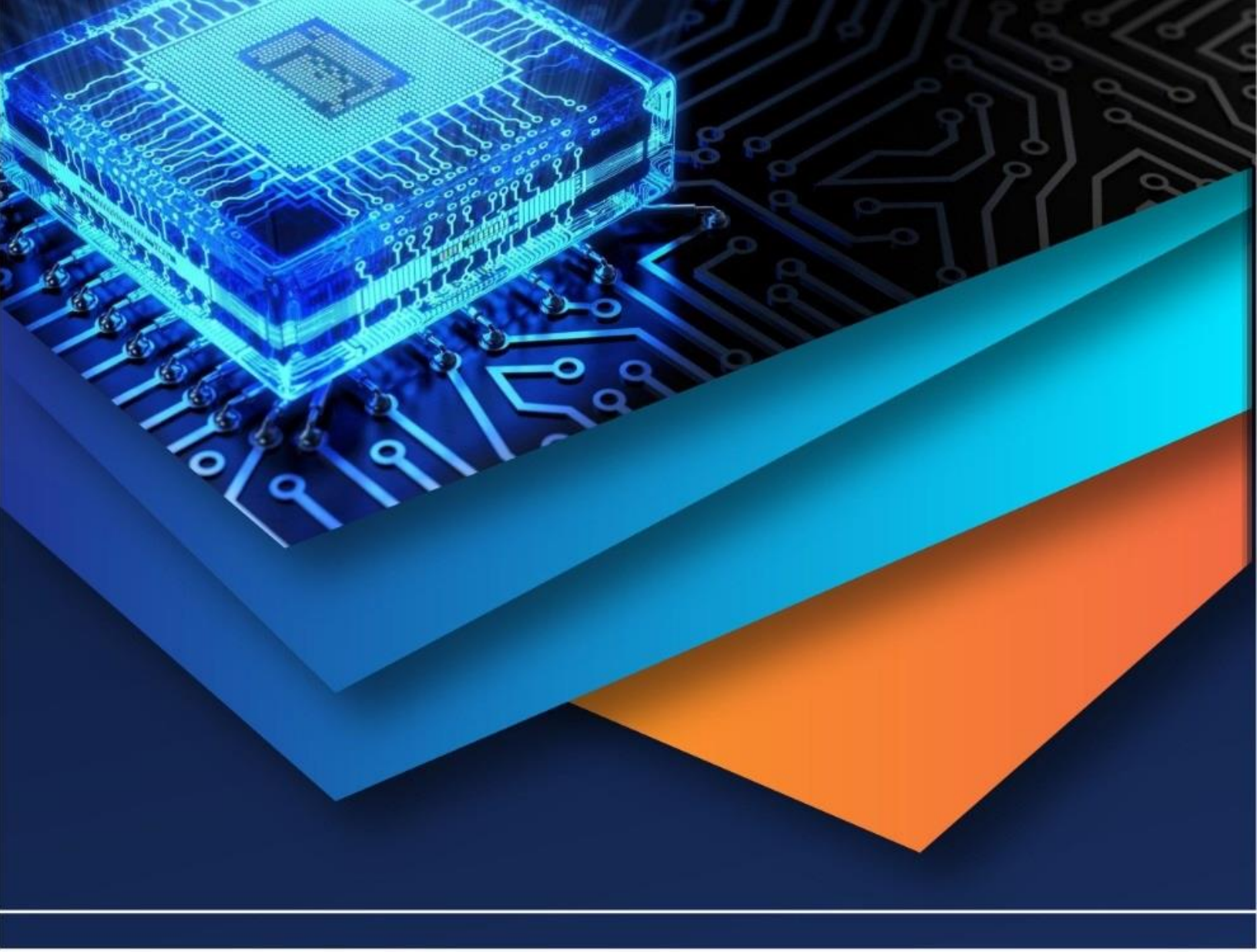
[1] N. K. Bhagat, Y. Vishnusai and G. N. Rathna, "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning," 2019 Digital Image Computing: Techniques and Applications (DICTA), 2019, pp. 1-8, Doi: 10.1109/DICTA47822.2019.8945850.

[2] M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917-126951, 2021, Doi: 10.1109/ACCESS.2021.3110912.

[3] Prachi Sharma, Radhey Shyam Anand, A comprehensive evaluation of deep models and optimizers for Indian sign language recognition, Graphics and Visual Computing, Volume 5,2021,200032,ISSN 2666-6294,https://doi.org/10.1016/j.gvc.2021.200032.

[4] (https://www.sciencedirect.com/science/article/pii/S2666629421000152)

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)