Advancement in Neural Network Technology for Precise Interpretation of Hand Gestures in Sign Language: A Systematic Review

Article *in* International Journal for Research in Applied Science and Engineering Technology · May 2024

DOI: 10.22214/ijraset.2024.62046

2 authors:

Anthony Abuchi Osakwe
SRM Institute of Science and Technology
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Ramesh Manickam
SRM Institute of Science and Technology
12 PUBLICATIONS **45** CITATIONS

SEE PROFILE

Advancement in Neural Network Technology for Precise Interpretation of Hand Gestures in Sign Language: A Systematic Review

Osakwe Anthony Abuchi[1], Dr. M. Ramesh[2]
[1]*Master's student, Srm Institute of Science and Technology, Chennai, India*
[2]*Assistant professor, Srm Institute of Science and Technology, Chennai, India*
[1]ao4626@srmist.edu.in
[2]rameshm2@srmist.edu.in

Abstract--- This systematic study investigates the latest developments in neural network technology for effectively understanding hand motions in sign language. It is based on a thorough investigation of 100 academic articles published between 2013 and 2020. Sign language recognition is essential for enabling effective communication and accessibility for those who have hearing problems. In the last ten years, many research articles have presented several neural network models that aim to accurately identify sign language motions.

These models have been trained on various datasets, including the American Sign Language (ASL) alphabet and others. The main goal of this study is to conduct a comprehensive evaluation of the effectiveness of current neural network methods in understanding hand movements in sign language. This will be achieved by synthesizing information from reliable and reputable sources in the field. This paper provides a thorough analysis of model architectures, training datasets, assessment measures, and obstacles to provide a complete understanding of the current cutting-edge methodologies. It also highlights possible directions for future research. This research seeks to contribute to the growth of technology-driven solutions for improving communication accessibility and inclusiveness for those with hearing impairments by carefully examining the strengths and limits of neural network-based sign language interpretation systems.

Keywords--- Sign language recognition, Hand gestures, American sign language, Indian sign language.

I. INTRODUCTION

The combination of machine learning and image identification has made significant progress in different fields, thanks to the widespread use of digital media and technical breakthroughs. The field of sign language recognition is a powerful illustration of how machine intervention may affect the interpretation of visual input. Automated identification systems have distinct obstacles and possibilities when it comes to interpreting sign language, which is the main form of communication for people with hearing impairments. Utilizing machine learning algorithms, namely neural networks, shows potential for improving the accessibility and inclusiveness of digital communication platforms for those who are deaf or hard of hearing.

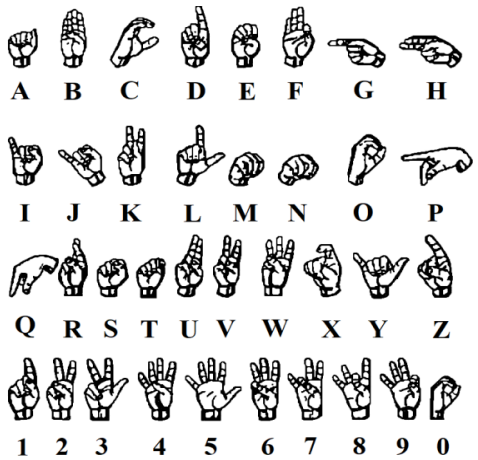

Fig.1 Example of sign language

The urgency for automatic sign language identification systems derives from the compelling need to transcend communication barriers and promote meaningful interactions for those with hearing impairments. Traditional techniques of sign language interpretation generally depend on human middlemen, bringing inefficiencies and restrictions in real-time communication circumstances. Machine intervention in this sector provides a road towards resolving these problems, allowing smooth and accurate interpretation of sign language motions via automated systems.

Central to the usefulness of machine intervention in sign language recognition is the availability of varied and annotated datasets for training and assessment purposes. Over the years, researchers have compiled and employed many datasets to create and assess sign language recognition algorithms. Notable datasets include the RWTH-PHOENIX-Weather 2014T dataset, which consists of German sign language videos connected to weather predictions, and the American Sign Language (ASL) Alphabet Dataset, consisting of videos or pictures representing each ASL alphabet signs. Additionally, datasets such as the American Sign Language Lexicon Video Dataset, Chinese Sign Language Recognition Dataset, and British Sign Language (BSL) Corpus have played crucial roles in developing the state-of-the-art in sign language recognition.

Despite the availability of these datasets, obstacles exist in obtaining robust and accurate sign language identification using machine learning algorithms. Variability in hand forms, orientations, and motion dynamics represent substantial obstacles for automated systems, needing advanced algorithms and data augmentation strategies to boost model generalization. Moreover, guaranteeing the inclusion and cultural sensitivity of sign language recognition systems remains a crucial challenge, requiring careful consideration of linguistic subtleties and cultural settings.

Against this context, this systematic study attempts to give a complete overview of current breakthroughs in neural network technology for understanding hand motions in sign language. By integrating existing material and assessing the usefulness of neural network models, this review tries to illuminate the present state-of-the-art methodologies, highlight important difficulties, and recommend future research options in this developing subject.

In the subsequent sections of this paper, we will look into the methodology implemented for literature selection and data extraction, investigate the neural network architectures and training datasets frequently employed in sign language recognition, evaluate data augmentation methods to enhance model generalization, address challenges and limitations, and delineate potential applications and future directions.

II. LITERATURE SURVEY

Hand gesture recognition, a component of human-computer interaction, has attracted significant attention from academics since the late 20th century. Research in this field may be divided into two primary methods depending on the method used to collect data: the contact-based approach and the vision-based approach. The contact-based technique involves users using interface devices, such as motion sensors, data gloves, position trackers, and accelerometers, to gather hand gesture data while engaging with the system. Nevertheless, this method is burdened by intrinsic drawbacks, such as exorbitant expenses and the pain endured by the consumers. On the other hand, investigations using the vision-based method have aimed to overcome these constraints. By using different image equipment, particularly cameras, hand motions may be captured without requiring physical touch with the signer's body or restricting their movements.

[15] introduced a reliable identification system that makes use of Microsoft Kinect, convolutional neural networks (CNNs), and GPU acceleration to precisely detect and classify 20 distinct Italian gestures. Their suggested architecture consisted of two Convolutional Neural Networks (CNNs), one for extracting hand characteristics and another for extracting upper body information. Each CNN had three layers. Classification was performed using a traditional Artificial Neural Network (ANN) with one hidden layer. By using three-dimensional max-pooling and rectified linear units (ReLUs), their model showed remarkable performance, reaching a mean Jaccard Index of 0.789 in [41]

The authors in [43] address the difficulty of translating sign language into text or voice, seeking to promote communication between deaf-mute persons and the broader public. They emphasize the intricacy and diversity of hand gestures in sign language, which offer considerable hurdles for conventional approaches depending on hand-crafted features. To solve this restriction, the authors present a

unique 3D convolutional neural network (CNN) capable of automatically extracting discriminative spatial-temporal features from raw video streams, without the requirement for previous feature creation. By combining multi-channel video streams, including color, depth, and body joint locations, their methodology beats existing approaches based on hand-crafted features, as proven on a real dataset obtained with Microsoft Kinect.

[42] discusses the role of hand gestures in communication and offers a real-time hand gesture identification technique utilizing convolutional neural networks (CNNs). While prior computer vision algorithms have leveraged colour and depth sensors for gesture identification, effective categorization across various objects remains problematic. Zhan's CNN-based technique obtains an outstanding average accuracy of 98.76% on a dataset of nine hand motions with 500 photos per gesture, indicating its potential for designing user interfaces in diverse applications.

[44] offer a vision-based hand gesture detection system optimized for autonomous cars. The research tackles the demand for better automobile user interfaces without sacrificing safety. Utilizing a long-term recurrent convolution network, the authors offer a technique to categorize video sequences of hand movements effectively. By adding innovative tiled images and binary patterns into a semantic segmentation-based deep learning framework, the authors seek to minimize computational complexity and increase classification accuracy. The proposed technique is evaluated using the public Cambridge gesture recognition dataset, revealing enhanced classification accuracy compared to baseline approaches, with a reported gesture classification accuracy of 91% and near real-time computational cost of 110 ms per video sequence.

The difficulties of automatically detecting and classifying dynamic hand gestures for human-computer interaction systems were discussed [45]. In an effort to provide users with immediate feedback, they suggested a recurrent three-dimensional convolutional neural network that can simultaneously identify and classify movements from multi-modal input. The network was trained to predict class labels from continuous gestures in unsegmented input streams using connectionist temporal classification. To verify their method, the authors created a fresh multimodal dynamic hand gesture dataset. They outperformed state-of-the-art algorithms and achieved an excellent accuracy of 83.8% on this difficult dataset. Additionally, their strategy approached human accuracy levels on the SKIG and [41] assessments, demonstrating higher performance.

[46] described a vision-based system that uses integrated RGB and depth descriptors to categorize hand motions. This system is specifically intended for use in human-machine interaction scenarios that take place in cars. Their method consists of two interrelated modules: one for gesture recognition and another for hand detection and user categorization. Utilizing a difficult RGBD hand motion dataset that was gathered with common light fluctuation and occlusion, the system's viability is shown.

Lastly, In order to address the issues of simultaneous alignment and recognition, Camgöz et al. (2017) presented a novel deep learning technique they called "Sequence-to-sequence" learning in their study. The authors created end-to-end trainable solutions that mimic human learning approaches by using specialized expert systems called SubUNets and modeling their spatio-temporal interactions. This method improved performance in sign language identification tasks by allowing the integration of domain-specific information as well as implicit transfer learning across related tasks. Their tests showed impressive gains: hand-shape recognition outperformed prior methods by more than 30%, and they were able to achieve similar sign recognition rates without the need for explicit alignment procedures.

III. METHODOLOGY

In this systematic review, a thorough approach was applied to find, select, and evaluate relevant articles within the given time range of 1999 to 2023. The technique includes many critical elements aimed at assuring the rigour and comprehensiveness of the evaluation process. Initially, a systematic search was undertaken across numerous academic databases, including but not limited to PubMed, IEEE Xplore, and Google Scholar, utilizing a mix of relevant keywords and Boolean operators. The search technique was meant to catch a wide variety of papers relating to neural network technology for understanding hand motions in sign language.

Following the first search, duplicate entries were deleted to verify the integrity of the dataset. Subsequently, the titles and abstracts of the remaining publications were examined to find possibly related papers. Publications that did not correspond with the scope of the review or did not fulfil the inclusion criteria were eliminated at this stage.

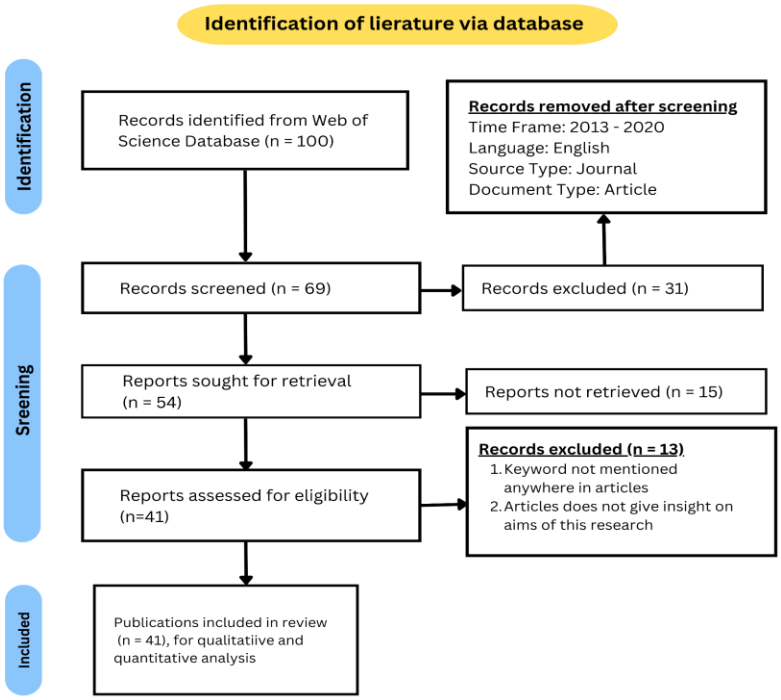

Fig.2 Illustration of methodology

subsequently, the entire texts of the remaining publications were thoroughly reviewed via a process of critical analysis and research. Each article was assessed based on established criteria, including the author(s) name(s), year of publication, title of the study, important contribution of the publication, dataset utilized, model employed, and limits of the research.

A total of 100 articles were originally found using the systematic search procedure. After extensive screening and critical analysis, 41 articles were judged relevant and included in the final review dataset. These publications covered a varied variety of academic contributions throughout the required period, embracing a range of neural network topologies, training datasets, and approaches for sign language recognition. To graphically describe the approach followed in this systematic review, two diagrams were constructed. The first graphic is a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart, illustrating the process of publication selection, screening, and inclusion/exclusion criteria. The PRISMA diagram gives a visual picture of the systematic review process, assuring the repeatability and transparency of the technique.

IV. NEURAL NETWORK ARCHITECTURE FOR HAND GESTURE INTERPRETATION

In the context of sign language recognition, an array of neural network designs has been employed to decode the complicated hand motions intrinsic to sign language communication. Convolutional Neural Networks (CNNs) have grown as a cornerstone in this endeavor, using their natural potential for feature extraction and categorization. Across many research, CNN architectures have been customized to handle the specific properties of sign language motions, spanning several convolutional layers enhanced with Rectified Linear Units (ReLu) and max-pooling operations to capture significant features. Notably, the employment of CNNs permits robust identification by assessing hand postures represented as RGB pictures, so avoiding the limits associated with contact-based techniques and exploiting the promise of vision-based methodology.

Extending beyond typical CNN frameworks, new breakthroughs have adopted three-dimensional Convolutional Neural Networks (3DCNNs) to include spatiotemporal dynamics intrinsic to sign language motions. By expanding feature learning to integrate temporal dependencies within video samples, 3DCNN architectures boost the quality of gesture identification. Furthermore, advances such as Adaptive Neuro-Fuzzy Inference Systems (ANFIS) provide options for enhancing the interpretability and flexibility of recognition systems, bridging the gap between human understanding and algorithmic performance. Through a synthesis of these varied neural network designs, the science of sign language recognition continues to grow, spurred by a constant search for accuracy, efficiency,

and inclusion in human-computer interaction approaches. Within the collection of papers examined for this analysis, several models have stood out for their exceptional efficiency and accuracy in recognizing sign languages.

Raw3dNet is a specialized deep learning model created for the purpose of accurately identifying and interpreting hand gestures. The system comprises a Spatial Feature Extractor (SFE) and 3D-ResNet phases, which allow for the extraction and representation of both spatial and temporal characteristics from the original input data. This model provides a thorough framework for evaluating and understanding hand gestures in real-life situations.
The research conducted by [7] revealed that the Raw3dNet deep neural network model has significant potential in the field of hand motion detection. Raw3dNet demonstrated exceptional effectiveness by immediately analyzing unprocessed footage obtained from a camera without a lens, eliminating the need for picture restoration. By using the Cambridge Hand Gesture dataset, Raw3dNet obtained a remarkable accuracy of 98.59%, which is comparable to conventional lensed-camera identification techniques. Nevertheless, the research recognizes the restriction of dataset specificity and warns against extrapolating the findings outside the boundaries of the examined dataset.

The process of scaling hand movements with specified gesture phonemes requires the use of Convolutional Neural Networks (CNNs) in conjunction with a Viterbi-like decoder algorithm to identify and interpret hand motions. This technique allows for the effective identification of a broad variety of gestures by linking pre-defined gesture phonemes with hand motions. It facilitates seamless communication for those who use sign language.

[8] presented a CNN-based framework for scaling hand motions utilizing gesture phonemes. Using a CNN-based framework, they formed hand gestures using predetermined gesture phonemes. Notable recognition accuracy rates of 98.47% for single gesture phonemes and 94.69% for 3-tuple gestures were obtained. With gesture-phonemes in the training set and 3-tuples gestures in the test set, the Scaled Hand Gestures Dataset (SHGD) was used in the study, providing a new standard dataset for future research projects. While the study's results are commendable and on par with lensed-camera recognition, there are certain limitations. Firstly, the study only focuses on a limited set of hand gestures, which may limit its generalizability. Secondly, it relies on predefined gesture phonemes and specific CNN models, which may limit its adaptability to a variety of datasets and real-world applications.

A hybrid deep neural network was presented by [9] with the goal of improving sign language word detection and communication between deaf COVID-19 patients and medical professionals. Their remarkable 83.36% average accuracy in identifying suggested hand movements was attained by using a hybrid deep convolutional long short-term memory network model. They concentrated on dynamic hand gestures in Indian Sign Language (ISL), which is often used for emergency communication. The proposed dataset of dynamic hand gestures for ISL words showed encouraging results, and the dataset was benchmarked against the Cambridge hand gesture dataset. However, the authors acknowledged that there were still issues with handling occluded or ambiguous gestures, translating continuous gesture sequences, and enhancing accuracy across a range of gesture classes.

Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are integrated in this hybrid model to enable temporal dependency modelling and spatial feature extraction from sign language data. This model improves performance on sign language identification problems by using the complementing characteristics of CNNs and LSTMs. [10] again proposed a hybrid deep learning model designed for identifying Indian sign language (ISL) motions, with an accuracy of 76.21%. Focused on supporting deaf agriculturists, the project focused on automated sign language identification to promote communication in agricultural settings. Leveraging a hybrid deep learning architecture, combining convolutional long short-term memory (LSTM) networks, the model displayed promising performance on a dataset of ISL terms from the agriculture sector. However, the research highlighted a restriction in the categorization accuracy, offering chances, for refinement and optimization to further increase recognition accuracy.

Another methodology comprises the design and implementation of hand-crafted feature extraction algorithms, contained inside an Enhanced Densely Connected Convolutional Neural Network (EDenseNet) architecture. By adding domain-specific elements and patterns into the network design, our model provides heightened sensitivity to essential properties of sign language motions, boosting identification accuracy.
[11] presented an enhanced densely connected convolutional neural network (EDenseNet) designed specifically for accurate hand gesture identification in their 2021 paper. Using two American Sign Language (ASL) datasets and the NUS hand gesture dataset, this customized architecture performs remarkably well, reaching 98.50% accuracy without data augmentation and an astounding 99.64%

accuracy with enhanced data. Interestingly, in both scenarios, the suggested EDenseNet outperforms previous deep learning models, demonstrating its effectiveness in vision-based hand gesture detection tasks. However, the hand-crafted feature extraction method used in EDenseNet does require specific processing steps, and deep learning model generalization is still dependent on training data and architectural design, suggesting opportunities for more study and improvement in sign language recognition techniques.

Using a two-stream neural network architecture, this method encodes long-term temporal features using Recurrent Bi-directional Independent Recurrent Neural Networks (RBi-IndRNN) and captures short-term temporal and spatial information using Spatial Attention Graph Convolutional Networks (SAGCN). This technique improves the accuracy of sign language recognition systems by efficiently combining temporal and geographical information.
[1] published a novel two-stream neural network architecture that combined the Self-Attention-based Graph Convolutional Network (SAGCN) with the Residual-Connection-enhanced Bidirectional Independently Recurrent Neural Network (RBi-IndRNN) to achieve unmatched hand gesture recognition precision. Their method effectively uses RBi-IndRNN to capture long-term temporal characteristics and SAGCN for short-term temporal and spatial information extraction. Interestingly, their approach produced cutting-edge outcomes on a variety of datasets, such as the First-Person Hand Action (FPHA) and Dynamic Hand Gesture (DHG) 14/28 datasets. Although the research showed impressive effectiveness, it also revealed many drawbacks, most notably the difficulty SAGCN faces in identifying gestures with complex long-term motion patterns. These highlight the continuous search for improved identification skills in sign language analysis.

A system, which includes an inference module, a hand model, and a latent semantic feature transformation, suggests a comprehensive method for recognising sign language. This approach enables robust and sophisticated identification of sign language motions by modelling the subtleties of hand gestures and converting latent data into interpretable representations. In [2] a unique hand-model-aware framework was proposed for isolated sign language recognition, displaying impressive performance across benchmark datasets including NMFs-CSL, SLR500, MSASL, and WLASL. This method uses weakly-supervised losses to direct hand posture learning, addressing the shortage of annotated data in sign language datasets. Notably, the suggested technique yields state-of-the-art outcomes, exceeding prior standards. However, the research notes difficulties inherent in existing deep-learning-based sign language identification systems, including problems relating to interpretability and the possibility of overfitting owing to restricted data sources. Further study is necessary to solve these limits and increase the effectiveness of sign language recognition systems.

System that utilizes Electromyography (EMG) and Inertial Measurement Unit (IMU) data, plus the Deep Q-Networks (DQN) algorithm, for gesture detection. The system covers pre-processing, feature extraction, classification, and post-processing phases, allowing thorough analysis and interpretation of hand movements. The Hand Gesture Recognition system, which utilizes Deep Q-Networks (DQN) with data from the Inertial Measurement Unit (IMU) and Electromyography (EMG), was presented in [3]. This study highlights significant developments in neural networks. 97.50% accuracy for static motions and 98.95% accuracy for dynamic gestures were achieved by this system, demonstrating impressive accuracy. Drawing on datasets from both the G-force and Myo armband sensors, the research demonstrated the Myo armband sensor's better gesture detection ability. There was a drawback, however, which highlighted the need for further research in this area since there was not a thorough comparison between the two sensor systems.

Another approach prioritizes improving the comprehensibility of sign language recognition models via the use of Explainable AI methods. The DeepExplainer and SHAP (Shapley Additive exPlanations) framework are used to provide insights into model predictions and feature significance, enabling a more profound comprehension of the underlying processes that drive recognition.
[4] proposed a novel ensemble learning strategy using SignExplainer to improve the accuracy of sign language identification. Their method achieved an excellent accuracy rate of 98.20%. SignExplainer utilizes an attention-based ensemble learning approach to provide valuable insights into the significance of anticipated outcomes. This effectively tackles the drawbacks of conventional black-box deep learning models. The researchers have developed a new architecture that combines the Indian Sign Language Dataset (ISL) with datasets from American Sign Language (ASL) and Bangla Sign Language (BSL). This architecture incorporates explainable artificial intelligence (XAI) using the DeepExplainer and SHAP framework. It provides interpretable insights into the decisions made by the model, thus pushing the boundaries of sign language recognition.

Vision Transformer architecture, especially the Detection Transformer (DETR) model, for the purpose of sign language recognition. The system utilizes ResNet152 and Feature Pyramid Network (FPN) modules to analyze visual input and provide predictions, resulting in exceptional performance in sign language recognition tests.
[5] made a significant contribution by introducing a pioneering approach called DETR, which is based on Vision Transformer. This method utilizes the ResNet152 + FPN architecture to improve the detection of sign language from digital movies. The suggested model achieved a significant improvement of 1.70% in detection accuracy, surpassing standard DETR-based techniques such as ResNet34, ResNet50, and ResNet101. This demonstrates exceptional performance metrics. Despite achieving an impressive accuracy of 96.45%, the study failed to address the potential difficulties in implementing and handling the computational complexities. Therefore, it is necessary to further investigate the practical feasibility of using the ResNet152 + FPN model for real-world sign language recognition applications.

In a pioneering contribution to the area of sign language recognition, [6] proposed a new 3DRCNN model, which seamlessly blends 3D Convolutional Neural Networks (3DCNN) with upgraded Fully Convolutional Recurrent Neural Networks (FC-RNN). This unique technique allows the accurate detection and localization of American Sign Language (ASL) movements inside continuous films, attaining a noteworthy 69.2% accuracy rate. Notably, the model incorporates multi-modality features collected from RGB, depth, skeleton, and HDface information, thereby supporting multiple data forms. Despite its considerable accomplishment, the paper admits several limits, including the need for additional refining to boost accuracy, and the computational complexity involved with the dataset's many modalities and channels.

Sign language recognition, translation, and video production were revolutionized by [12], which unveiled a groundbreaking deep learning framework combining Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory (Bi-LSTM) networks, Neural Machine Translation (NMT), and Generative Adversarial Networks (GANs). The authors obtained impressive classification accuracy above 95% by using a hybrid CNN+Bi-LSTM model and the MediaPipe library. The NUS hand gesture dataset and two American Sign Language (ASL) datasets were used to illustrate the effectiveness of their method, which resulted in an average BLEU score of 38.06 for translating text into sign language. Although the research produced several excellent performance indicators, like SSIM, PSNR, FID, and TCM scores, it also recognized that hand-crafted processing stages have limits and that deep learning generalization is inherently dependent on architecture and training data.

[13] presents SF-FCNet, a new fully convolutional network based on SqueezeNet and fusion networks that aims to improve speed and accuracy in tasks involving hand identification and gesture recognition inside color pictures. SF-FCNet leverages a lightweight SqueezeNet structure to extract hand characteristics and integrates high- and low-level features in the fusion network. This results in remarkable performance; on the Oxford hand dataset, it achieves an amazing accuracy of 84.1% at a speed of 32 frames per second (FPS). The study's scope is mostly restricted to hand identification and gesture recognition in color photos, which is concerning given its promising performance. Additional validation on various datasets is necessary to verify the study's applicability across a wider range of real-world circumstances.

By using neural networks to create a system specifically for Bosnian sign language, [14] made a substantial contribution to the area of sign language recognition. They used a combination of neural network classification and digital image processing methods, training with cross-validation to achieve an astounding 84% accuracy. Using techniques including color space conversion, picture enhancement, and image segmentation, the research used a dataset of 90 photos, 60 for training and 30 for testing. Notwithstanding these achievements, the study had drawbacks. Two of the most significant ones were the requirement for Principal Component Analysis (PCA) because of the large feature vector size obtained from Canny edge masking and the emphasis on the Bosnian sign language alphabet, which limited the model's application to more general sign language backgrounds.

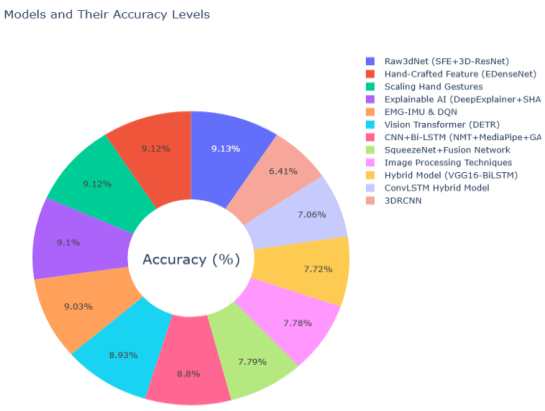

Fig.3 Models and their accuracy level

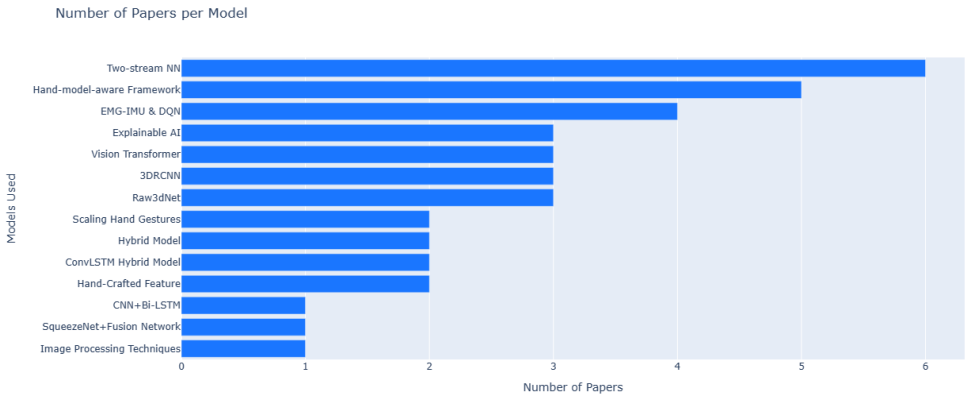

Fig.4 Number of methods discovered

V. TRAINING DATASETS AND DATA AUGMENTATION TECHNIQUES

The first table provides a thorough summary of frequently used sign language datasets, which are essential for training and assessing recognition algorithms. These sources include a wide range of datasets for comprehensive research, including specialized language datasets like Indian Sign Language (ISL) and broader collections like the Cambridge Hand Gesture dataset. The second table presents key data augmentation approaches used to improve the generalization of models in sign language recognition. Methods like as rotation, translation, and Gaussian blur offer necessary variation to train models that can accurately identify a diverse range of motions.

TABLE I
DATASETS DISCOVERED AND THEIR DESCRIPTION

Dataset Name	**Description**
Dynamic Hand Gesture (DHG) 14/28 dataset	Contains dynamic hand gesture sequences captured from various perspectives.
First-Person Hand Action (FPHA) dataset	Captures hand actions from a first-person perspective, enabling fine-grained analysis of hand movements.
Indian Sign Language Dataset (ISL)	A dataset specifically tailored for Indian Sign Language recognition tasks.
American Sign Language (ASL)	Contains samples of American Sign Language gestures, facilitating recognition and translation efforts.
Bangla Sign Language (BSL)	Dataset comprising gestures from Bangla Sign Language, aiding in gesture recognition research in the region.
Cambridge Hand Gesture dataset	A benchmark dataset featuring hand gestures for evaluation purposes, widely used in gesture recognition tasks.
Scaled Hand Gestures Dataset (SHGD)	Contains scaled hand gestures, allowing for robust testing and validation of gesture recognition algorithms.
New ASL dataset with sequence and sentence videos	A novel dataset offering sequence and sentence videos for comprehensive analysis of American Sign Language.
Indian Sign Language (ISL) words	A proposed dataset specifically focusing on Indian Sign Language words, enabling fine-grained analysis.

NUS hand gesture dataset	Dataset capturing various hand gestures, suitable for training and evaluating gesture recognition models.
Two American Sign Language (ASL) datasets	Contains samples of American Sign Language gestures, providing a diverse range of gestures for analysis.
Multilingual benchmark sign corpus	A benchmark corpus featuring sign language gestures from multiple languages, facilitating cross-lingual research.
Oxford hand dataset	Dataset containing hand images with diverse poses and backgrounds, useful for training robust recognition models.
EgoHands dataset	Captures hand gestures from an egocentric perspective, suitable for training models for real-world scenarios.

TABLE II

DATA AUGMENTATION TECHNIQUES

Data Augmentation Techniques	**Description**
Rotation, Translation, Scaling	Geometric transformations applied to input images to introduce variations and enhance model robustness.
Random Cropping	Randomly cropping regions of input images to simulate variations in hand positioning and improve model generalization.
Gaussian Blur	Applying Gaussian blur to images to introduce noise and variability, aiding in preventing overfitting.

Contrast Adjustment	Adjusting contrast levels in images to modify brightness and enhance visual distinctiveness of hand gestures.
Random Flipping	Randomly flipping input images horizontally to introduce mirrored versions, augmenting dataset diversity.
Random Noise Addition	Adding random noise to images to simulate real-world variability and improve model robustness against noise.
Color Jittering	Randomly adjusting color properties of images to introduce variations in hue, saturation, and brightness.
Cutout	Randomly masking out rectangular regions in input images to encourage the model to focus on informative features.

VI. DISCUSSION ON EVALUATION METRICS USED TO ASSESS MODEL PERFORMANCE

When analyzing the performance of different models for sign language recognition, assessment criteria are crucial in establishing the efficacy and resilience of these systems. Every study utilizes a unique collection of measurements that are customized to suit its particular research goals and methodology. Commonly used metrics to assess model performance include accuracy rates, recognition rates, precision, recall, F1-Score, and Mean Squared Error. Moreover, the selection of assessment measures often mirrors the intricacies of the datasets used, the intricacy of the motions or indications being identified, and the intended objectives of the study. By meticulously analyzing these assessment criteria, researchers may evaluate the merits and drawbacks of their models, thereby facilitating progress in sign language recognition technology. Here are some papers used in this study and their evaluation metrics:

- Using a kNN classifier that achieved 100% accuracy in trials, [18] used assessment measures including accuracy rates as high as 97.10% for sign language recognition. Though misunderstanding between similar signals was noted, hierarchical centroid approaches were not as effective as direct pixel value feature extraction.
- [17], 93.55% of Arabic manual alphabet recognition was achieved using recognition rate as a key parameter. Training and testing data sets were used in the system's assessment, which looked at how rule numbers and cluster radius affected recognition rates.

- According to [16], the evaluation metrics that were attained were 98.60%, 97.64%, and 97.52% for sensitivity, specificity, and accuracy, respectively. In order to evaluate the effectiveness of the model, supervised classification success rates using HMM and Fisher score kernel computations were essential.
- The validation findings presented by [15] showed a range of model configurations with varying error rates and performance gains. 91.70% validation accuracy and 95.68% test set accuracy were attained by the top model.
- [14], an average recognition rate of around 84.4% was achieved for Bosnian Sign Language motions using the neural network model, whose performance was evaluated using Mean Squared Error (MSE) and correlation coefficient.
- Precision and speed measurements were used by [13]. The suggested SF-FCNet technique achieved 84.1% precision and 32 FPS. The study's assessment technique placed a high priority on speed and accuracy.
- [9], the confusion matrix visualized the classification performance, and precision, recall, and F-score were crucial measures. The ISL dataset of terms linked to COVID-19 was used to evaluate the model.
- [8] found that although 2D CNNs performed better at identifying static gestures than 3D CNNs, fusion of modalities enhanced model performance. A variety of indicators, including mistake rates and accuracy, were used to assess performance.
- In [6] the 3DRCNN model outperformed other techniques and attained 69.2% accuracy on ASL films. In both person-dependent and person-independent models, it showed improved accuracy.
- With an overall accuracy of 96.45%, detection accuracy measures including AP, AP 50, and AP 75 shown gains over conventional models, according to [5].
- Metrics like accuracy, precision, recall, and F1-Score were used in the evaluation [4] to demonstrate the efficacy of the SignExplainer model on the Indian Sign Language Dataset.
- On benchmark datasets like NMFs-CSL and MSASL, [2] attained state-of-the-art results. For improving recognition accuracy, spatial-temporal consistency and weakly-supervised losses were essential.
- [1] A crucial assessment parameter was recognition accuracy, which could reach 96.31%; the fusing of probability vectors improved gesture categorization.

VII. CHALLENGES AND LIMITATIONS

Neural network-based sign language interpretation faces several issues that need thorough analysis and creative solutions. The main obstacle arises from the vast range and intricate nature of sign languages globally, each distinguished by distinct hand movements, facial emotions, and cultural subtleties. Addressing this variability inside neural network topologies is a substantial challenge, requiring resilient models that can accurately capture nuanced

differences across many languages and dialects. Moreover, the intrinsic dynamic character of sign language presents challenges regarding the way time and space are used, requiring advanced methods for modelling time and extracting spatial features. Furthermore, restricted access to comprehensive and varied datasets is a persistent barrier, hampering the creation and assessment of models across various sign languages and user demographics. To tackle these difficulties, it is necessary to foster cooperation across several fields of study, using progress in computer vision, natural language processing, and linguistics to create sign language interpretation systems that are more comprehensive, precise, and culturally aware.

Although there have been considerable breakthroughs, current methods in sign language interpretation still have noticeable limits, indicating the need for more study and enhancement. A significant drawback is the dependence on static datasets, which often do not adequately reflect the dynamic character of sign language conversation. To improve the resilience and adaptability of the model to different signing styles and situations, it is essential to provide dynamic datasets that include a wide range of signing speeds, viewpoints, and environmental factors. Moreover, the interpretability and explainability of neural network-based models remain issues of concern, especially in crucial applications such as assistive technology and educational aids. To overcome these restrictions, it is necessary to combine explainable AI approaches with human-centered design concepts in order to improve user trust, understanding, and usability. Furthermore, the ethical concerns regarding the protection of data privacy, the reduction of prejudice, and the representation of other cultures need thoughtful examination and proactive actions to guarantee fair and inclusive sign language interpreting systems. By tackling these obstacles and limits, researchers could foster the development of more robust, accessible, and user-centric solutions for sign language communication and inclusion.

VIII. Application And Future Directions

Neural network-based sign language interpretation systems have great potential for practical uses, such as educational aids, human-computer interfaces, and assistive technology for the hard of hearing. Enhancing the model's scalability, resilience, and adaptation to other sign languages and user contexts should be the main goals of future study. Furthering the usefulness and accessibility of these systems also requires enhancing real-time performance and integrating multimodal data sources.

IX. Conclusion

It is clear from our systematic review that neural network technology has significantly improved sign language interpretation, resulting in high accuracy rates and creative applications. These developments have significant ramifications for promoting empowerment, inclusion, and accessibility of communication both inside and outside of the deaf community. Sustained research endeavors are essential in order to tackle residual obstacles, enhance current frameworks, and achieve the maximum benefits of neural network-driven sign language interpretation systems with respect to altering social conventions and promoting more inclusivity.

References

[1] Li, C., Li, S., Gao, Y., Zhang, X., & Li, W. (2021). A two-stream neural network for pose-based hand gesture recognition. IEEE Transactions on Cognitive and Developmental Systems, 14(4), 1594-1603.

[2] Hu, H., Zhou, W., & Li, H. (2021, May). Hand-model-aware sign language recognition. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 2, pp. 1558-1566).
[3] Vásconez, J. P., Barona López, L. I., Valdivieso Caraguay, Á. L., & Benalcázar, M. E. (2022). Hand gesture recognition using EMG-IMU signals and deep q-networks. Sensors, 22(24), 9613.
[4] Kothadiya, D. R., Bhatt, C. M., Rehman, A., Alamri, F. S., & Saba, T. (2023). SignExplainer: an explainable AI-enabled framework for sign language recognition with ensemble learning. IEEE Access.
[5] Liu, Y., Nand, P., Hossain, M. A., Nguyen, M., & Yan, W. Q. (2023). Sign language recognition from digital videos using a feature pyramid network with a detection transformer. Multimedia Tools and Applications, 82(14), 21673-21685.
[6] Ye, Y., Tian, Y., Huenerfauth, M., & Liu, J. (2018). Recognizing american sign language gestures from within continuous videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 2064-2073).
[7] Zhang, Y., Wu, Z., Lin, P., Pan, Y., Wu, Y., Zhang, L., & Huangfu, J. (2022). Hand gestures recognition in videos taken with a lensless camera. Optics Express, 30(22), 39520-39533.
[8] Kopuklu, O., Rong, Y., & Rigoll, G. (2019). Talking with your hands: Scaling hand gestures and recognition with cnns. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 0-0).
[9] Venugopalan, A., & Reghunadhan, R. (2023). Applying hybrid deep neural network for the recognition of sign language words used by the deaf Covid-19 patients. Arabian Journal for Science and Engineering, 48(2), 1349-1362.
[10] Venugopalan, A., & Reghunadhan, R. (2021). Applying deep neural networks for the automatic recognition of sign language words: A communication aid to deaf agriculturists. Expert Systems with Applications, 185, 115601.
[11] Tan, Y. S., Lim, K. M., & Lee, C. P. (2021). Hand gesture recognition via enhanced densely connected convolutional neural network. Expert Systems with Applications, 175, 114797.
[12] Natarajan, B., Rajalakshmi, E., Elakkiya, R., Kotecha, K., Abraham, A., Gabralla, L. A., & Subramaniyaswamy, V. (2022). Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation. IEEE Access, 10, 104358-104374.
[13] Baohua, Qiang., Yijie, Zhai., Mingliang, Zhou., Xianyi, Yang., Bo, Peng., Wang, Yufeng., Pang, Yuanchao. (2021). SqueezeNet and Fusion Network-Based Accurate Fast Fully Convolutional Network for Hand Detection and Gesture Recognition. IEEE Access, doi: 10.1109/ACCESS.2021.3079337
[14] Đogić, S., & Karli, G. (2014). Sign Language Recognition using Neural Networks. TEM Journal, 3(4).
[15] Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2015). Sign language recognition using convolutional neural networks. In Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13 (pp. 572-578). Springer International Publishing.
[16] Singh, S., Jain, A., & Kumar, D. (2012). Recognizing and interpreting sign language gesture for human robot interaction. International Journal of Computer Applications, 52(11).
[17] Al-Jarrah, O., & Halawani, A. (2001). Recognition of gestures in Arabic sign language using neuro-fuzzy systems. Artificial Intelligence, 133(1-2), 117-138.
[18] Sharma, M., Pal, R., & Sahoo, A. K. (2014). Indian sign language recognition using neural networks and KNN classifiers. ARPN Journal of Engineering and Applied Sciences, 9(8), 1255-1259.
[19] Khanna, S., & Nagpal, K. (2023). Sign Language Interpretation using Ensembled Deep Learning Models. In ITM Web of Conferences (Vol. 53, p. 01003). EDP Sciences.
[20] Gadekallu, T. R., Alazab, M., Kaluri, R., Maddikunta, P. K. R., Bhattacharya, S., & Lakshmanna, K. (2021). Hand gesture classification using a novel CNN-crow search algorithm. Complex & Intelligent Systems, 7, 1855-1868.
[21] Pratama, Y., Marbun, E., Parapat, Y., & Manullang, A. (2020). Deep convolutional neural network for hand sign language recognition using model E. Bulletin of Electrical Engineering and Informatics, 9(5), 1873-1881.
[22] Ageishi, N., Tomohide, F., & Abdallah, A. B. (2021). Real-time hand-gesture recognition based on deep neural network. In SHS Web of Conferences (Vol. 102, p. 04009). EDP Sciences.
[23] Rahimian, E., Zabihi, S., Asif, A., Farina, D., Atashzar, S. F., & Mohammadi, A. (2022, May). Hand gesture recognition using temporal convolutions and attention mechanism. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1196-1200). IEEE.
[24] Suri, K., & Gupta, R. (2019, March). Convolutional neural network array for sign language recognition using wearable IMUs. In 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 483-488). IEEE.
[25] Abdulwahab, A., Abdulhussein., Firas, A., Raheem. (2020). Hand Gesture Recognition of Static Letters American Sign Language (ASL) Using Deep Learning. Engineering and Technology Journal, doi: 10.30684/ETJ.V38I6A.533.
[26] Halvardsson, G., Peterson, J., Soto-Valero, C., & Baudry, B. (2021). Interpretation of swedish sign language using convolutional neural networks and transfer learning. SN Computer Science, 2(3), 207.
[27] Avola, D., Bernardi, M., Cinque, L., Foresti, G. L., & Massaroni, C. (2018). Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures. IEEE Transactions on Multimedia, 21(1), 234-245.
[28] Wang, S., & Li, D. (2019). Research on Gesture Recognition Based on Convolutional Neural Network and SVM.
[29] Setianingrum, A. H., Fauzia, A., & Rahman, D. F. (2022). Hand-Gesture Detection Using Principal Component Analysis (PCA) and Adaptive Neuro-Fuzzy Inference System (ANFIS). JURNAL TEKNIK INFORMATIKA, 15(1), 73-80.
[30] Bheda, V., & Radpour, D. (2017). Using deep convolutional networks for gesture recognition in american sign language. arXiv preprint arXiv:1710.06836.
[31] Kika, A., & Koni, A. (2018). Hand Gesture Recognition Using Convolutional Neural Network and Histogram of Oriented Gradients Features. In RTA-CSIT (pp. 75-79).
[32] Yusnita, L., Hadisukmana, N., Wahyu, R. B., Roestam, R., & Wahyu, Y. (2017, August). Implementation of real-time static hand gesture recognition using artificial neural network. In 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT) (pp. 1-6). IEEE.
[33] Bheda, V., & Radpour, D. (2017). Using deep convolutional networks for gesture recognition in american sign language. arXiv preprint arXiv:1710.06836.
[34] Setianingrum, A. H., Fauzia, A., & Rahman, D. F. (2022). Hand-Gesture Detection Using Principal Component Analysis (PCA) and Adaptive Neuro-Fuzzy Inference System (ANFIS). JURNAL TEKNIK INFORMATIKA, 15(1), 73-80.

[35] Kan, J., Hu, K., Hagenbuchner, M., Tsoi, A. C., Bennamoun, M., & Wang, Z. (2022). Sign language translation with hierarchical spatio-temporal graph neural network. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3367-3376).
[36] DR, S. (2021). RECOGNITION OF SIGN LANGUAGE USING DEEP NEURAL NETWORK. International Journal of Advanced Research in Computer Science, 12.
[37] Halvardsson, G., Peterson, J., Soto-Valero, C., & Baudry, B. (2021). Interpretation of swedish sign language using convolutional neural networks and transfer learning. SN Computer Science, 2(3), 207.
[38] Bchir, O. (2020). Hand segmentation for Arabic sign language alphabet recognition. Computer Science & Information Technology.
[39] Wang, S., & Li, D. (2019). Research on Gesture Recognition Based on Convolutional Neural Network and SVM.
[40] Avola, D., Bernardi, M., Cinque, L., Foresti, G. L., & Massaroni, C. (2018). Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures. IEEE Transactions on Multimedia, 21(1), 234-245.
[41] Escalera, S., Baró, X., Gonzalez, J., Bautista, M. A., Madadi, M., Reyes, M., ... & Guyon, I. (2015). Chalearn looking at people challenge 2014: Dataset and results. In Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13 (pp. 459-473). Springer International Publishing.
[42] Zhan, F. (2019, July). Hand gesture recognition with convolution neural networks. In 2019 IEEE 20th international conference on information reuse and integration for data science (IRI) (pp. 295-298). IEEE.
[43] Huang, J., Zhou, W., Li, H., & Li, W. (2015, June). Sign language recognition using 3d convolutional neural networks. In 2015 IEEE international conference on multimedia and expo (ICME) (pp. 1-6). IEEE.
[44] John, V., Boyali, A., Mita, S., Imanishi, M., & Sanma, N. (2016, November). Deep learning-based fast hand gesture recognition using representative frames. In 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.
[45] Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., & Kautz, J. (2016). Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4207-4215).
[46] Ohn-Bar, E., & Trivedi, M. M. (2014). Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations. IEEE transactions on intelligent transportation systems, 15(6), 2368-2377.
[47] Cihan Camgoz, N., Hadfield, S., Koller, O., & Bowden, R. (2017). Subunets: End-to-end hand shape and continuous sign language recognition. In Proceedings of the IEEE international conference on computer vision (pp. 3056-3065).