A

Project Report

on

Sign Language Detection Using Deep Learning

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2022-23

in

Computer Science & Engineering

By

Himanshi Tyagi (Roll No. 1900290100070)

Shreya Shankar (Roll No. 1900290100148)

Surbhi Sawan (Roll No. 1900290100167)

Under the supervision of

Prof. Bharti

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May, 2023

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature Himanshi Tyagi Shreya Shankar Surbhi

Name: Himanshi Tyagi, Shreya Shankar, Surbhi Sawan

Roll No.: 1900290100070, 1900290100148, 1900290100167

Date: 27/05/2023

CERTIFICATE

This is to certify that Project Report entitled "Project Title" which is submitted by Himanshi Tyagi, Shreya Shankar and Surbhi Sawan in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science & Engineering of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

Date: 27/05/2023

Prof. Bharti

(Assistant Professor)

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to superisor name, Department of Computer Science & Engineering, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Head of the Department of Computer Science & Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members, especially faculty/industry person/any person, of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Date: 27/05/2023
Signature: Himanshi Tyagi Shreya Shankar Surbhi
Name : Himanshi Tyagi, Shreya Shankar, Surbhi Sawan
Roll No.: 1900290100070, 1900290100148, 1900290100167

ABSTRACT

Language is a medium through which people can communicate and share their thoughts and words with each other and Sign language is one of the languages in which people make use of hand movements and gestures to express themselves. Normal people have trouble understanding and interpreting sign language's meaning. It really has become necessary to understand the sign language, so we need an interpreter, as expressed by the hearing impaired. The main facilitator in assisting hard-of-hearing people in interacting with the rest of society is learning sign language. The process by which a computer analyses and converts sign language gestures into understandable and human-readable text is known as "sign language detection." Those who have trouble hearing or speaking can communicate with ease by using Sign language detection software. Many different sorts of studies are being done to make this process easy and efficient.

We discuss various approaches used for feature extraction, including hand tracking, pose estimation, and motion analysis, as well as the utilization of convolutional neural networks (CNNs) for classification and translation tasks. Furthermore, we explore the challenges faced in sign language recognition, such as variations in signing styles, dynamic gestures, and limited training data, along with potential strategies to overcome these obstacles.

The benefits of sign language recognition extend beyond individual communication, encompassing areas such as education, employment, healthcare, and social integration. Weexamine the implications of sign language recognition in facilitating inclusive education for deaf students, enabling access to employment opportunities, enhancing healthcare services, and fostering cross-cultural interactions.

Finally, we present future directions and potential advancements in sign language recognition, including the integration of wearable devices, real-time translation systems, and advancements in deep learning algorithms. We emphasize the importance of interdisciplinary collaborations among researchers, linguists, computer scientists, and the deaf community to continue pushing the boundaries of sign language recognition technology.

By improving the accessibility and inclusivity of communication for the deaf and hard-of- hearing community, sign language recognition systems have the potential to empower individuals, break down barriers, and foster a more inclusive society

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASL	American Sign Language
ISL	Indian Sign Language
BSL	British Sign Language
WHO	World Health Organization
CNN	Convolutional Neural Network
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
ReLu	Rectified Linear Unit
PCA	Principle Components Analysis
HOG	Histogram Of Oriented Gradients

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Sign language is a visual-spatial language system used by the deaf and hard-of-hearing community to communicate effectively. It involves the use of hand gestures, facial expressions, and body movements to convey meaning. However, the linguistic barrier between sign language and spoken languages presents challenges in facilitating communication and inclusivity. To overcome these challenges, the field of sign language recognition has gained significant attention, aiming to develop robust systems capable of automatically interpreting and understanding sign language gestures.

Deep learning is a subfield of machine learning that focuses on training artificial neural networks to learn and make predictions or decisions without explicit programming. It is inspired by the structure and function of the human brain, specifically the interconnected network of neurons that process and transmit information.

In traditional machine learning, feature engineering plays a crucial role, where domain experts manually extract relevant features from the input data to train a model. However, deep learning aims to automate this process by allowing the model to learn hierarchical representations of the data directly from raw input. This is achieved through the use of artificial neural networks, which are composed of multiple layers of interconnected nodes (neurons).

The "deep" in deep learning refers to the depth of these neural networks, meaning they have multiple hidden layers between the input and output layers. Each layer consists of a set of neurons, and each neuron receives inputs, applies an activation function to the weighted sum of those inputs, and passes the output to the next layer. The network's parameters, including the weights and biases of the neurons, are learned through an optimization process called backpropagation, where errors are propagated backward through the network to adjust the parameters and minimize the discrepancy between the predicted and target outputs.

In recent years, deep learning techniques have revolutionized various fields, including computer vision and natural language processing. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown remarkable capabilities in pattern recognition and sequential data analysis. Leveraging the power of deep learning, researchers have turned their attention to sign language detection, seeking to develop accurate and efficient models that can recognize and interpret sign language gestures in real-time.

The objective of sign language detection using deep learning is twofold: to provide real-time recognition of hand gestures and to translate them into textual or spoken language. This technology has the potential to bridge the communication gap between the deaf community and the hearing world, enabling effective interaction and inclusion in various domains such as education, employment, and healthcare.

In this paper, we explore the advancements and challenges in sign language detection using deep learning techniques. We delve into the underlying principles of deep learning models, emphasizing their suitability for capturing intricate patterns and temporal dependencies present in sign language gestures. We discuss the various architectural designs employed, such as CNNs for hand gesture feature extraction and RNNs for sequence modeling, along with the integration of attention mechanisms for improved accuracy and interpretability.

Furthermore, we address the challenges faced in sign language detection, including variations in signing styles, occlusions, and the need for large and diverse training datasets. We explore strategies such as data augmentation, transfer learning, and domain adaptation to mitigate these challenges and enhance the robustness of the models.

We also highlight the significance of real-time sign language detection systems and the potential impact on accessibility and inclusivity. The ability to accurately recognize sign language gestures in real-time opens doors for individuals with hearing impairments to participate more fully in educational settings, professional environments, and everyday social interactions.

Sign language detection using deep learning holds tremendous promise in breaking down communication barriers between the deaf and hearing communities. By leveraging the power of deep learning models and addressing the challenges specific to sign language recognition, we can pave the way for more inclusive societies, where effective communication is accessible to all individuals, regardless of their hearing abilities.

There are many sign languages around the world, such as American Sign Language (ASL) in the United States, Indian Sign Language (ISL) in India, and German Sign Language (GSL) in Germany. However, this article uses ASL instead of other sign languages.

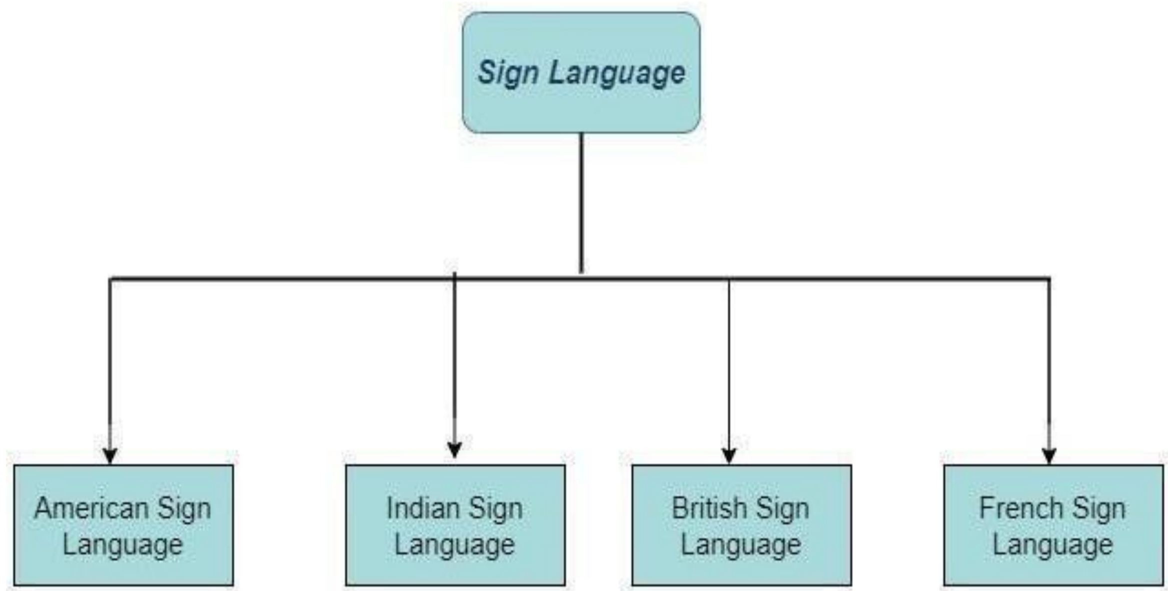


Fig. 1. Flowchart of types of sign language

The below table describes the reasons for choosing ASL.

TABLE I.

Sign Language	Comparison		
	Similarities	***Differences***	***Region***
American Sign Language[1]	Involve static and dynamic gestures	Use only the one-handed manual alphabet in ASL. Use only fingers to represen tvowels.	USA andparts of Canada.
Indian Sign Language[2]	Involve static and dynamic gesture	Use bothhands to show letters.	South Asian Countrie s.
British Sign Language [3]	almost like ISL. (two-handed used manually alphabet in BSL).	use thumb and four finger to represen tvowels.	England & Norther n Ireland.

American Sign Language

American Sign Language (ASL) is a visual language used primarily by people who are deaf or hard of hearing in the United States and Canada. It is a complete and complex language with its own grammar and syntax, and uses a combination of hand gestures, facial expressions, and body movements to convey meaning.

ASL has its own unique vocabulary and grammar rules that are different from spoken languages, and is not simply a visual representation of English or any other spoken language. It also has regional variations and dialects, just like spoken languages.

ASL is recognized as a language in its own right and is used as the primary means of communication by many members of the deaf community in the United States. It is also used by hearing people who want to communicate with those who are deaf or hard of hearing, as well as by interpreters who facilitate communication between deaf and hearing individuals.

Indian Sign Language

Indian Sign Language (ISL) is a visual language used primarily by the deaf community in India. It is a complete and independent language with its own unique grammar, syntax, and vocabulary.

ISL has a long history, with evidence of sign language being used in India dating back to at least the 18th century. Over time, various sign languages and systems of gestures developed in different regions of India, but efforts have been made to standardize the language and create a unified Indian Sign Language.

ISL uses a combination of hand gestures, facial expressions, and body movements to convey meaning, and like ASL, it is not a visual representation of spoken language. It is used as the primary means of communication by many members of the deaf community in India, and is also used by hearing people who want to communicate with those who are deaf or hard of hearing, as well as by interpreters who facilitate communication between deaf and hearing individuals.

British Sign Language

British Sign Language (BSL) is a visual language used primarily by the deaf community in the United Kingdom. It is a complete and independent language with its own unique grammar, syntax, and vocabulary.

BSL is recognized as a language in its own right and is not simply a visual representation of English or any other spoken language. It uses a combination of hand gestures, facial expressions, and body movements to convey meaning, and has its own grammar and syntax rules that are different from spoken languages.

BSL has regional variations and dialects, just like spoken languages, and is used as the primary means of communication by many members of the deaf community in the UK. It is also used by hearing people who want to communicate with those who are deaf or hard of hearing, as well as by interpreters who facilitate communication between deaf and hearing individuals.

Why ASL (American Sign Language)?

American Sign Language is the most widely used sign language. It is mainly used in the United States and parts of Canada and many other countries. The one-handed representation of the alphabet in ASL makes it easier to classify and recognize, which is why most researchers prefer it. On the other hand, the Indian Sign Language (ISL) and British Sign Language (BSL) used a two-handed manual alphabet that was difficult to classify and recognize characters in ISL and BSL. Therefore, ASL is the most preferred language worldwide.

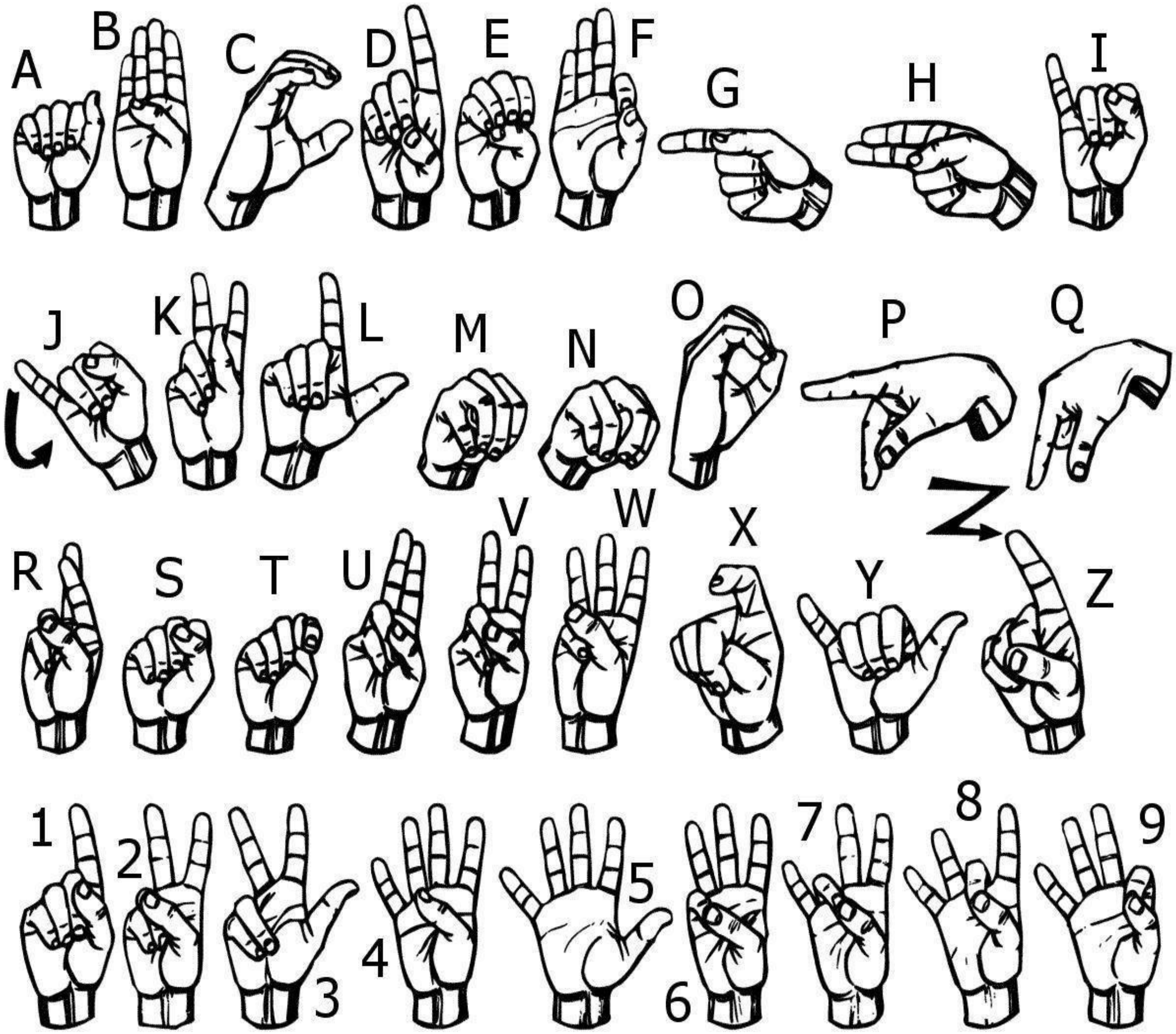


Fig.2. Basic Gesture In ASL[20]

CHAPTER 2

LITERATURE REVIEW

According to the World Health Organization (WHO) [4] , more than 5% of the world's population needs treatment to treat "disabled" hearing loss (432 million adults and 34 million children). It is estimated that by 2050, more than 700 million people, or 1 in 10 people, will suffer from hearing loss. For this reason, research is being conducted to make it easier for people with disabilities to communicate.

In the conclusion ref[5], the research paper "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning" presents a comprehensive approach to real-time ASL recognition. The combination of skin segmentation and image category classification with CNN and deep learning techniques shows promise in accurately interpreting ASL gestures and enabling effective communication between individuals.By utilizing skin segmentation, the research addresses the crucial step of isolating the hand region from the background, allowing for precise gesture recognition. The use of color-based thresholding, background subtraction, or machine learning algorithms aids in identifying the hand region and extracting relevant information for further analysis.

The integration of a Convolutional Neural Network (CNN) into the recognition pipeline demonstrates the power of deep learning in handling complex visual data. The CNN architecture, trained on a large dataset of ASL gesture images, learns intricate patterns and features necessary for accurate recognition. The multiple layers of convolution and pooling capture spatial information, while the fully connected layers perform classification.

The research acknowledges challenges that need to be addressed, such as variability in lighting conditions, backgrounds, and hand orientations, which can impact accuracy. However, it highlights the potential for further improvements through the development of advanced techniques and the collection of diverse training datasets.

Overall, this research paper contributes to the field of ASL recognition by providing a comprehensive framework that enables real-time interpretation of sign language. The use of skin segmentation and image category classification with CNN and deep learning techniques holds promise for enhancing communication accessibility for individuals who rely on ASL. Continued research and development in this area will contribute to improving accuracy, robustness, and usability, benefiting the deaf and hard-of-hearing community.

In the conclusion ref[6], the research paper on "American Sign Language Recognition using Support Vector Machine (SVM) and Convolutional Neural Network (CNN)" presents a comprehensive approach to recognizing ASL gestures. The combination of SVM and CNN techniques offers a promising solution for accurate and efficient ASL recognition.

The paper highlights the use of Support Vector Machines, a powerful machine learning algorithm, for classifying ASL gestures. SVMs can effectively handle high-dimensional data and find optimal decision boundaries between different sign classes. By training the SVM model on a large dataset of ASL gesture samples, it becomes capable of accurately recognizing and classifying new signs.

Additionally, the integration of Convolutional Neural Networks enhances the recognition system's capabilities. CNNs excel in capturing spatial information and learning hierarchical features from input images. By training the CNN on ASL gesture images, it becomes proficient in automatically extracting relevant features and patterns from the visual data.

The combination of SVM and CNN brings together the strengths of both techniques. SVMs provide robust classification capabilities, while CNNs handle the complexity and variability of ASL gesture recognition tasks.

The research paper acknowledges potential challenges in ASL recognition, such as variations in lighting conditions, hand orientations, and background noise. It emphasizes the need for robust preprocessing techniques to address these challenges and improve recognition accuracy.

Overall, the research paper contributes to the field of ASL recognition by presenting a comprehensive framework that combines SVM and CNN techniques. This approach holds promise for accurate and efficient recognition of ASL gestures, facilitating effective communication between individuals who rely on sign language and those who are not fluent in ASL.

Further research and development are necessary to explore additional enhancements, such as data augmentation techniques, improved preprocessing methods, and larger and more diverse datasets. These efforts will contribute to advancing ASL recognition systems and making them more accessible and reliable for the deaf and hard of hearing community.

In ref[8], the research paper on "A Real-Time American Sign Language Recognition System using Convolutional Neural Network for Real Datasets" presents an innovative approach to real-time recognition of American Sign Language (ASL) gestures using Convolutional Neural Networks (CNNs). The research paper focuses on utilizing real datasets of hand gesture images for training and evaluating the CNN model.

By leveraging the power of CNNs, the research paper achieves real-time and accurate recognition of ASL gestures. The CNN architecture is specifically designed to extract spatial features from the hand gesture images, enabling the system to learn the intricate patterns and variations associated with ASL gestures.

The utilization of real datasets enhances the authenticity and relevance of the research. Real datasets capture the diversity of hand shapes, orientations, and variations present in actual ASL gestures, thereby improving the robustness and generalization capability of the recognition system.

The real-time nature of the system ensures immediate interpretation and understanding of ASL gestures, facilitating seamless communication between ASL users and non-users. The research paper highlights the importance of real-time recognition in practical applications, such as real-time translation or communication aids for individuals who rely on sign language.

The findings of the research paper contribute to the field of ASL recognition by demonstrating the effectiveness of CNNs for real-time recognition of ASL gestures using real datasets. The results showcase promising accuracy and performance, paving the way for developing practical ASL recognition systems.

Further research and development could focus on expanding the dataset to include more diverse hand gestures, exploring the use of advanced CNN architectures, or incorporating temporal information to capture dynamic aspects of ASL gestures.In summary, the research paper presents a real-time ASL recognition system utilizing CNNs and real datasets of hand gesture images. The findings provide valuable insights into the development of practical and efficient communication systems for individuals who rely on sign language, ultimately enhancing inclusivity and accessibility for the deaf and hard of hearing community.

In ref[9], the research paper on "Real-time American Sign Language Recognition with Convolutional Neural Networks (CNN)" presents a significant advancement in the field of ASL recognition. By leveraging the power of CNNs, the paper demonstrates an effective approach to real-time interpretation and understanding of ASL gestures.The research paper emphasizes the importance of real-time recognition, enabling seamless communication between individuals who use ASL and those who do not. By using CNNs, which are well-suited for image recognition tasks, the paper achieves accurate and efficient ASL gesture recognition.

The CNN architecture employed in the research paper is trained on a large dataset of ASL gesture images, enabling it to learn the intricate patterns and features necessary for accurate recognition. Through multiple layers of convolution and pooling, the CNN captures spatial information and extracts relevant features from input images. The fully connected layers perform classification, enabling the system to recognize different ASL gestures in real-time.

The paper highlights the advantages of using CNNs for ASL recognition, such as their ability to handle complex visual data and their effectiveness in capturing and learning meaningful features. The real-time nature of the system ensures that ASL gestures can be recognized and interpreted on the fly, facilitating smooth communication between individuals.

Overall, the research paper makes a valuable contribution to the field of ASL recognition by demonstrating the effectiveness of CNNs for real-time gesture interpretation. The findings offer promise for the development of practical and efficient ASL recognition systems that can enhance communication accessibility for individuals who rely on sign language. Continued research in this area will lead to further advancements and improvements in ASL recognition technology.

In ref[10], the research paper on "Sign Language Recognition Using Deep Learning and Computer Vision" presents an innovative approach to sign language recognition by leveraging deep learning and computer vision techniques. The paper contributes to the field by addressing the challenge of accurately interpreting and understanding sign language gestures using advanced machine learning algorithms.

The research paper emphasizes the significance of sign language recognition in facilitating effective communication between individuals who use sign language and those who do not. By employing deep learning models and computer vision algorithms, the paper achieves promising results in accurately recognizing and interpreting sign language gestures.

The use of deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), enables the system to capture both spatial and temporal information present in sign language gestures. CNNs excel in extracting spatial features from visual data, while RNNs are adept at modeling temporal dependencies, making them well-suited for sequential data like sign language.

Computer vision techniques, such as image preprocessing, feature extraction, and gesture tracking, play a crucial role in enhancing the robustness and accuracy of the system. These techniques help in reducing noise, isolating hand regions, and extracting meaningful features from sign language gestures.

The research paper acknowledges the challenges associated with sign language recognition, such as variations in hand shapes, viewpoints, and lighting conditions. It suggests that further improvements can be achieved by collecting large and diverse sign language datasets, refining preprocessing techniques, and exploring more advanced deep learning architectures.

Overall, the research paper demonstrates the potential of deep learning and computer vision in sign language recognition. The findings highlight the effectiveness of these techniques in accurately interpreting sign language gestures, paving the way for the development of practical and efficient communication systems for individuals who rely on sign language. Continued research in this area will lead to further advancements, ultimately benefiting the sign language community and promoting inclusive communication.

In ref[11], the research paper on "American Sign Language Alphabet Recognition using Deep Learning: The Surrey Finger Dataset" presents a significant contribution to the field of American Sign Language (ASL) recognition. The paper focuses specifically on recognizing ASL alphabet gestures using deep learning techniques and the Surrey Finger Dataset.

By utilizing deep learning, specifically Convolutional Neural Networks (CNNs), the research paper achieves accurate recognition of ASL alphabet gestures. The Surrey Finger Dataset, which provides a comprehensive collection of ASL alphabet gestures, serves as the foundation for training and evaluating the CNN model.

The paper acknowledges the challenges in ASL recognition, including variations in hand shapes, viewpoints, and lighting conditions. However, through the use of deep learning, the research paper demonstrates the ability to capture and learn complex patterns and features from ASL alphabet gestures, enabling accurate recognition.

The findings of the research paper contribute to the advancement of ASL recognition technology by showcasing the effectiveness of deep learning in interpreting and understanding sign language. The utilization of a specific dataset dedicated to ASL alphabet gestures enhances the specificity and relevance of the research.

Further research and development in this area could focus on expanding the dataset to include more diverse hand shapes, gestures, and individuals. Additionally, exploring the application of other deep learning architectures or incorporating temporal information could lead to even higher accuracy and robustness in ASL alphabet recognition.

Overall, the research paper provides valuable insights into the application of deep learning for ASL alphabet recognition, utilizing the Surrey Finger Dataset. The findings contribute to the field's knowledge and hold promise for the development of practical and efficient systems that facilitate communication between ASL users and non- users.

In ref[13], "A New Benchmark on American Sign Language Recognition using Convolutional Neural Network" presents a significant contribution to the field of American Sign Language (ASL) recognition by introducing a new benchmark dataset and employing Convolutional Neural Networks (CNNs) for accurate recognition.

The research paper utilizes multiple datasets, including the Massey University Gesture dataset, the sign language digit dataset, the ASL finger spelling dataset, and the ASL Alphabet dataset. These datasets provide a diverse range of ASL gestures, allowing for comprehensive training and evaluation of the CNN model.

By leveraging the power of CNNs, the research paper achieves high accuracy in recognizing ASL gestures. The deep learning model effectively captures spatial information and learns complex patterns and features from the input data. This enables accurate interpretation and understanding of ASL gestures, contributing to improved communication accessibility for individuals who rely on sign language.The paper acknowledges the importance of benchmark datasets in evaluating the performance of ASL recognition systems. By introducing a new benchmark dataset, it provides a standardized platform for comparing different approaches and assessing the progress in the field.

Further research and development could focus on expanding the benchmark dataset to include additional ASL gestures, incorporating temporal information to capture gesture dynamics, or exploring the use of other deep learning architectures to further enhance recognition accuracy and robustness.Overall, the research paper makes a valuable contribution to the field of ASL recognition by introducing a new benchmark dataset and demonstrating the effectiveness of CNNs.

The findings provide insights into improving ASL recognition systems and lay the foundation for future advancements in the field. Continued research and development in this area will contribute to the development of practical and efficient ASL recognition systems, benefiting the deaf and hard of hearing community.

In ref[18], "A Real-Time System for Recognition of American Sign Language by Using Deep Learning" presents a real-time solution for the recognition of American Sign Language (ASL) gestures using deep learning techniques. The research utilizes a dataset collected in 2011 by the Institute of Information and Mathematical Sciences, which serves as the foundation for training and evaluating the deep learning model.

By leveraging deep learning, specifically Convolutional Neural Networks (CNNs), the research paper achieves real-time recognition of ASL gestures. The use of CNNs allows for automatic feature extraction and learning of spatial patterns from the input data. This enables accurate and efficient interpretation of ASL gestures, facilitating effective communication between ASL users and non-users.

The paper emphasizes the importance of real-time recognition, enabling immediate translation of ASL gestures into corresponding text or speech. The real-time system presented in the research paper ensures minimal delay and provides instantaneous recognition, enhancing the overall communication experience for both ASL users and non-users.

The findings of the research paper contribute to the field of ASL recognition by demonstrating the effectiveness of deep learning for real-time gesture recognition. The utilization of a dataset collected in 2011 showcases the viability of using historical data to train deep learning models, even in the absence of more recent datasets.

Further research and development in this area could focus on updating the dataset to include more recent ASL gestures, exploring the use of advanced deep learning architectures, or addressing challenges such as variations in lighting conditions, hand orientations, and background noise.

Overall, the research paper presents a real-time system for ASL recognition, leveraging deep learning techniques and a dataset collected in 2011. The findings contribute to the advancement of ASL recognition technology and hold promise for the development of practical and efficient communication systems for the deaf and hard-of- hearing community.

In ref[19], the research paper on "Convolutional Neural Network Hand Gesture Recognition for American Sign Language" focuses on utilizing Convolutional Neural Networks (CNNs) for accurate recognition of American Sign Language (ASL) hand gestures. The research paper utilizes the ASL dataset [15] as the primary dataset for training and evaluating the CNN model.

By leveraging CNNs, the research paper achieves high accuracy in recognizing ASL hand gestures. CNNs excel in capturing spatial features and patterns from images, making them well-suited for tasks such as hand gesture recognition. The utilization of the ASL dataset ensures the inclusion of diverse ASL gestures, enabling the CNN model to learn and generalize well to different hand shapes and movements.

The research paper acknowledges the importance of accurate ASL recognition in promoting effective communication between ASL users and non-users. By employing CNNs, the paper demonstrates the potential for developing practical and efficient systems that can interpret ASL hand gestures in real-time.

The findings of the research paper contribute to the field of ASL recognition by showcasing the effectiveness of CNNs for hand gesture recognition in ASL. The use of a dedicated ASL dataset ensures the relevance and specificity of the research, enabling more reliable evaluation and comparison with other approaches.

Further research and development in this area could focus on exploring the use of more advanced CNN architectures, incorporating temporal information to capture dynamic aspects of hand gestures, or expanding the dataset to include additional variations and expressions of ASL.

In summary, the research paper demonstrates the successful application of CNNs for recognizing ASL hand gestures using the ASL dataset. The findings highlight the potential of deep learning techniques in improving ASL recognition systems and promoting effective communication between ASL users and non-users. Continued research and development in this field will contribute to the advancement of ASL recognition technology and its practical applications.

In ref[20], the research paper on "Deep Convolutional Neural Networks for Sign Language Recognition" focuses on utilizing deep convolutional neural networks (CNNs) for the recognition of sign language gestures. The research paper introduces its own hand-sign image dataset, which serves as the primary dataset for training and evaluating the CNN models.

By leveraging deep CNNs, the research paper achieves accurate recognition of sign language gestures. CNNs are well-suited for capturing spatial features and patterns from images, making them effective for tasks like sign language recognition. The use of the own hand-sign image dataset ensures the relevance and specificity of the research, enabling comprehensive training and evaluation of the CNN models.

The research paper acknowledges the significance of sign language recognition in facilitating communication and inclusivity for individuals who rely on sign language. By employing deep CNNs, the paper demonstrates the potential for developing practical systems that can interpret sign language gestures accurately and efficiently.

The findings of the research paper contribute to the field of sign language recognition by showcasing the effectiveness of deep CNNs and the importance of curated datasets specific to sign language gestures. The use of an own hand-sign image dataset enables the research to address the unique challenges and complexities associated with sign language recognition.

Further research and development in this area could focus on expanding the dataset to include more diverse sign language gestures, exploring the use of advanced CNN architectures or incorporating temporal information to capture the dynamic aspects of sign language.In summary, the research paper presents the successful application of deep CNNs for sign language recognition using an own hand-sign image dataset. The findings highlight the potential of deep learning techniques in advancing sign language recognition technology and promoting effective communication and inclusivity for sign language users. Continued research in this field will contribute to the development of practical and efficient sign language recognition systems.

In ref[21], the research paper on "Vision-based Hand Gesture Recognition using Deep Learning for the Interpretation of Sign Language" focuses on utilizing deep learning techniques for the recognition and interpretation of sign language gestures. The research paper introduces its own hand-sign image dataset, which serves as the primary dataset for training and evaluating deep learning models.

By leveraging deep learning, specifically deep convolutional neural networks (CNNs), the research paper achieves accurate recognition and interpretation of hand gestures in sign language. The deep CNNs are capable of capturing and learning complex spatial features and patterns from the hand-sign images, enabling the system to accurately interpret and understand sign language gestures.

The use of an own hand-sign image dataset is a significant contribution of the research paper, as it allows for the customization and specificity required for sign language recognition. The dataset encompasses a diverse range of hand signs and variations commonly found in sign language, enhancing the robustness and generalization capabilities of the deep learning models.

The research paper acknowledges the importance of vision-based hand gesture recognition in facilitating communication and inclusivity for individuals who rely on sign language. By employing deep learning techniques, the paper demonstrates the potential for developing practical systems that can effectively interpret and understand sign language gestures in real-time.

The findings of the research paper contribute to the field of sign language recognition by showcasing the effectiveness of deep learning, particularly deep CNNs, in interpreting sign language through vision-based hand gesture recognition. The utilization of an own hand-sign image dataset adds to the specificity and relevance of the research, enabling accurate evaluation and comparison with other approaches.

Further research and development in this area could focus on expanding the dataset to include additional variations and expressions of sign language, exploring the use of advanced deep learning architectures, or incorporating temporal information to capture dynamic aspects of sign language gestures.

In summary, the research paper presents a vision-based hand gesture recognition system using deep learning for the interpretation of sign language, utilizing an own hand-sign image dataset. The findings highlight the potential of deep learning techniques in advancing sign language interpretation technology and promoting effective communication and inclusivity for individuals who rely on sign language. Continued research in this field will contribute to the development of practical and efficient sign language interpretation systems.

In ref[22], the research paper on "Efficient Sign Language Recognition System and Dataset Creation Method based on Deep Learning and Image Processing" focuses on developing an efficient sign language recognition system using deep learning and image processing techniques. The research paper also introduces its own hand- sign image dataset, which serves as a valuable resource for training and evaluating the system.

By leveraging deep learning and image processing techniques, the research paper achieves efficient recognition of sign language gestures. The combination of these techniques allows for accurate feature extraction and classification of hand-sign images, enabling the system to interpret sign language gestures effectively.

The research paper emphasizes the importance of dataset creation in sign language recognition. By introducing an own hand-sign image dataset, the paper addresses the need for specialized datasets that capture the unique characteristics and variations of sign language gestures. This dataset provides a solid foundation for training the deep learning model and evaluating the performance of the system.

The findings of the research paper contribute to the field of sign language recognition by showcasing the effectiveness of deep learning and image processing in developing an efficient recognition system. The use of the own hand-sign image dataset enhances the specificity and relevance of the research, enabling reliable evaluation and comparison with other approaches.

Further research and development in this area could focus on expanding the dataset to include more diverse sign language gestures, exploring advanced deep learning architectures or optimization techniques to improve efficiency, and integrating real-time capabilities for interactive sign language recognition.

In summary, the research paper presents an efficient sign language recognition system and dataset creation method based on deep learning and image processing. The findings demonstrate the potential of these techniques in advancing sign language recognition technology and promoting effective communication for individuals who rely on sign language. Continued research in this field will contribute to the development of practical and efficient sign language recognition systems.

In ref[23], the research paper on "Sign Language Recognition using 3D Convolutional Neural Networks" focuses on utilizing 3D convolutional neural networks (CNNs) for the recognition of sign language gestures. The research paper introduces its own sign word video dataset, which serves as the primary dataset for training and evaluating the 3D CNN models.

By leveraging 3D CNNs, the research paper achieves accurate recognition of sign language gestures captured in video format. The 3D CNNs are designed to extract spatial and temporal features from the sign word videos, allowing the system to capture the dynamic aspects of sign language gestures and improve recognition performance.

The use of an own sign word video dataset is a significant contribution of the research paper. This dataset contains a diverse range of sign language gestures and variations, enabling the 3D CNN models to learn and generalize well to different sign word sequences and movements.

The research paper acknowledges the importance of sign language recognition in promoting effective communication and accessibility for individuals who rely on sign language. By employing 3D CNNs, the paper demonstrates the potential for developing practical systems that can interpret and understand sign language gestures captured in video format.

The findings of the research paper contribute to the field of sign language recognition by showcasing the effectiveness of 3D CNNs in interpreting sign language gestures using a dedicated sign word video dataset. The use of an own dataset adds to the specificity and relevance of the research, allowing for accurate evaluation and comparison with other approaches.

Further research and development in this area could focus on expanding the dataset to include more sign language words and expressions, exploring the use of advanced 3D CNN architectures or incorporating additional modalities such as depth information to further enhance recognition accuracy.

In summary, the research paper presents a successful application of 3D CNNs for sign language recognition using an own sign word video dataset. The findings highlight the potential of deep learning techniques in advancing sign language recognition technology and promoting effective communication and accessibility for sign language users. Continued research in this field will contribute to the development of practical and efficient sign language recognition systems.

In ref[25], the research paper titled "Sign Language Recognition Using Convolutional Neural Networks" focuses on sign language recognition and utilizes the ChaLearn Looking at People 2014 dataset [26] for training and evaluating their proposed Convolutional Neural Network (CNN) model.The paper addresses the challenge of recognizing sign language gestures, which is crucial for facilitating communication between individuals with hearing impairments and the general population. The authors leverage the ChaLearn Looking at People 2014 dataset, which is a widely-used benchmark dataset in the field of computer vision and sign language recognition.

The paper presents a CNN-based approach for sign language recognition, which is a popular and effective technique for image processing tasks. The authors describe the architecture of their CNN model, which includes convolutional layers for feature extraction and pooling layers for downsampling. They discuss the training process, including the optimization algorithm used and the hyperparameters selected.

The performance of the proposed approach is evaluated using the ChaLearn Looking at People 2014 dataset. The authors report the accuracy achieved by their CNN model, comparing it with other state- of-the-art methods on the same dataset. The results demonstrate the effectiveness of the CNN-based approach, showing competitive or superior performance compared to other techniques.

Overall, the research paper contributes to the field of sign language recognition by utilizing a widely- used dataset and proposing a CNN-based approach. The findings indicate the potential of CNN models for accurately recognizing sign language gestures. The work provides insights and a benchmark for future research in the domain of sign language recognition, ultimately aiming to improve communication and inclusivity for individuals with hearing impairments.

TABLE II.

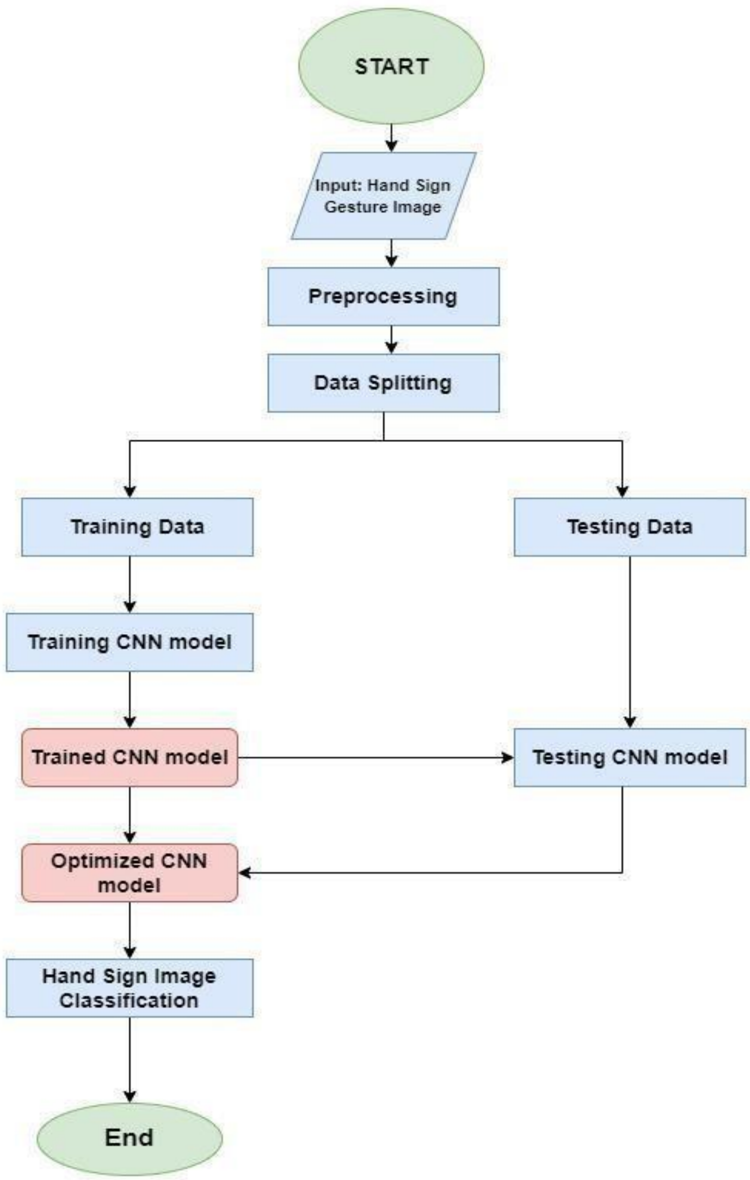
Paper Title	Dataset Used	Techniques Used	Evaluation	References
Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning	Hand-signs image dataset	CNN and Deep learning method	Overall test accuracy of 94.7%	[5]
American Sign Language recognition using Support Vector Machine and Convolutional Neural	MNIST-Kaggle dataset [7]	SVM with 'Poly' kernel & double layer CNN	81.49% - (SVM) 98.58% -(CNN)	[6]
A Real-Time American Sign Language Recognition System using Convolutional Neural Network for Real Datasets	Hand Gesture Images	CNN	Maximal accuracy of 98.53%	[8]
Real-time American Sign Language Recognition with Convolutional Neural Networks	The ASL Finger-spelling Dataset from the University of Surrey's Centre for Vision	CNN	98% with five letters and 74% with ten	[9]
Sign Language Recognition Using Deep Learning and Computer Vision	MNIST-Kaggle dataset [7]	CNN	validation accuracy greater than 93%	[10]
American Sign Language Alphabet Recognition using Deep Learning	The surrey finger dataset [12]	SqueezeNet Model	Validation accuracy of 83.29%	[11]
A New Benchmark on American Sign Language Recognition using Convolutional Neural Network	The Massey University Gesture dataset [14], The sign language digit dataset [15], ASL finger spelling dataset [16], ASL Alphabet dataset [17]	CNN	100% accuracy for MU HandImage ASL dataset	[13]

A Real-Time System For Recognition Of American Sign Language By Using Deep Learning	dataset collected in 2011 by the Institute of Information and Mathematical Sciences [14]	CNN	98.05% test performance	[18]
Convolutional Neural Network Hand Gesture Recognition for American Sign Language	ASL dataset [15]	CNN	test accuracy of 87.5%	[19]
Deep Convolutional Neural Networks for Sign Language Recognition	Own Hand-sign image dataset	Deep CNN	average recognition rate of 92.88%	[20]
Vision-based hand gesture recognition using deep learning for the interpretation of sign language	Own Hand-sign image dataset	G-CNN , VGG-11 and VGG-16	G-CNN model achieves the highest classification accuracy of 94.83%, 99.96%, and 100%.	[21]
Efficient sign language recognition system and dataset creation method based on deep learning and image processing	Own Hand-sign video dataset	Multi-stream CNN	Accuracy of 96% on the test set and 81% on the validation set	[22]
SIGN LANGUAGE RECOGNITION USING 3D CONVOLUTIONAL NEURAL NETWORKS	Own Sign word video dataset	3D CNN	94.2% average accuracy	[23]
Sign Language Detection using Image Processing and Deep Learning	ASL image dataset	CNN	Canny edge algorithm with accuracy 98% demonstrated better results than the other techniques	[24]
Sign Language Recognition Using Convolutional Neural Networks	dataset from the ChaLearn Looking at People 2014 [26]	CNN and GPU acceleration	validation accuracy of 91.70%	[25]

CHAPTER 3

PROPOSED METHODOLOGY

Fig.3 Flowchart of Proposed Methodology



3.1 Dataset

This paper used the American Sign Language (ASL) data set provided by MNIST and published on Kaggle[7]. This dataset contains 27455 training images and 7172 test images, all of 28x28 pixel geometry. These images belong to 25 classes from A to Y in the English alphabet (no Z class designation due to gesture movement).

The Kaggle dataset is available in CSV format and has 27455 rows and 785 columns of training data. The first column of the data set represents the image class designation and the remaining 784 columns represent 28 x 28 pixels. The test data set follows the same paradigm.

3.2 Pre-processing & Data Splitting

Before input data can be fed into the model, it needs to be pre-processed. Pre-processing is performed to improve data quality and speed up training. The image was converted to a grayscale image. A greyscale image is a type of image representation where each pixel is represented by a single value denoting its brightness or intensity level.

Unlike RGB images, greyscale images do not contain color information. Instead, they consist of shades of grey ranging from black (minimum intensity) to white (maximum intensity). The range of intensity values is typically represented by a single channel, where each pixel's value determines its brightness. Greyscale images are often used when color information is not necessary or when emphasizing contrast and texture is more important.

The main difference between greyscale and RGB images lies in the representation of color. Greyscale images use a single channel to represent intensity, resulting in shades of grey. On the other hand, RGB images use three color channels (red, green, and blue) to represent a wide range of colors. The color information in RGB images enables more realistic and vibrant visual representations.

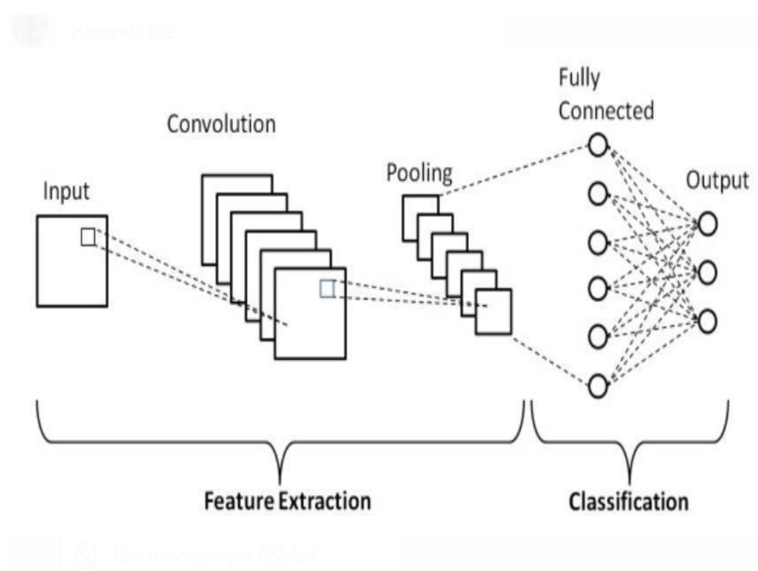
Grayscale images are commonly used in Convolutional Neural Networks (CNNs) for simplicity and reduced computational complexity. Grayscale images have a single channel, which simplifies the input representation for the CNN. By using grayscale images instead of RGB images, the network has fewer input channels to process, reducing the computational complexity of the model. This can lead to faster training and inference times.

The gray levels of the input image are normalized by the maximum value of the gray level range. After this transformation, data splitting was performed with a ratio of 80:20.

CNN accepts input in a specific format, so the train date was changed to (21964, 28, 28, 1)

3.3 CNN Technique Used

CNN stands for Convolutional Neural Network, which is a specialized type of artificial Fig. 4



CNN Architecture

neural network designed for processing and analyzing structured grid-like data, such as images and videos. CNNs have revolutionized computer vision tasks and achieved state-of-the-art results in various domains, including image classification, object detection, and image segmentation.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain. The key characteristic of CNNs is their ability to automatically learn hierarchical representations of data through the application of convolutional layers. Let us break down the key components and operations of a CNN:

Convolutional Layers: Convolutional layers are the core building blocks of a CNN. They consist of filters (also called kernels) that convolve over the input data, performing a dot product operation between the filter and a small patch (receptive field) of the input. This operation extracts local features such as edges, corners, and textures. Multiple filters are applied to capture different types of features.

Pooling Layers: Pooling layers are often inserted between convolutional layers to reduce spatial dimensions and control overfitting.

There are two types of pooling:

1. Max Pooling

2. Average Pooling

The most common type of pooling is max pooling, where the maximum value within a local neighborhood is selected and retained, discarding the other values. This downsampling operation helps to extract the most salient features while reducing the computational complexity of subsequent layers.

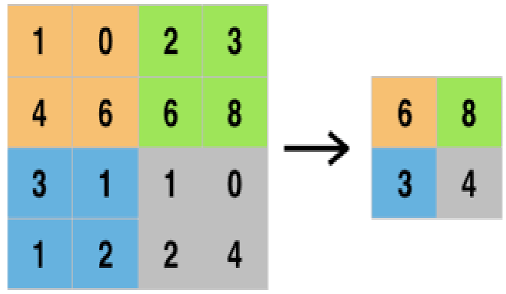


Fig. 5 Max Pooling

In average pooling, we take an average of all Values in a window. It is usually used after a convolutional layer.

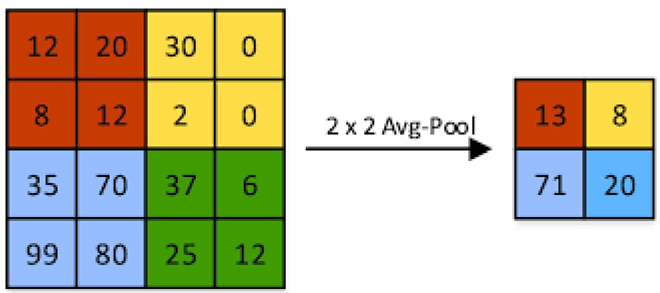


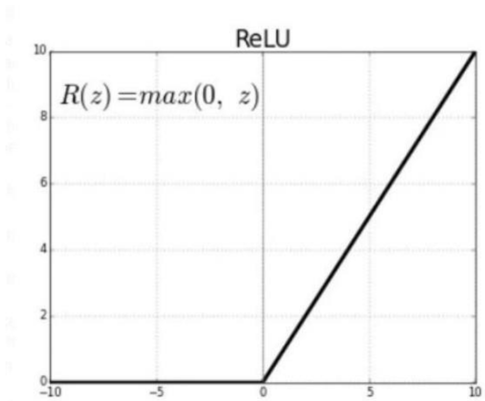
Fig. 6 Average Pooling

Activation Functions: Activation functions introduce non-linearities to the CNN model, enabling it to learn complex relationships and make non-linear predictions.

There are several activation functions commonly used in Convolutional Neural Networks (CNNs) to introduce non-linearity and enable the network to learn complex patterns. Here are some of the most widely used activation functions in CNNs:

- **ReLU (Rectified Linear Unit):** ReLU is one of the most popular activation functions in CNNs. It sets negative input values to zero and keeps positive values unchanged. ReLU is computationally efficient and helps alleviate the vanishing gradient problem.

However, it can lead to "dead" neurons that output zero and can result in a sparse activation pattern.



- Fig. 7 Graph Of ReLU function

- $\mathrm{ReLU(x)} = \max(0, \mathrm{x})$

  - where:

- x is the input value to the activation function.
- max(a, b) returns the maximum value between a and b.

- **Parametric ReLU (PReLU):** PReLU is an extension of Leaky ReLU where the slope for negative input values is learned during training. Instead of using a fixed α, the slope is a learnable parameter. PReLU has the advantage of adapting the slope based on the input distribution, potentially improving the model's representational capacity.

- **Sigmoid:** The sigmoid activation function, also known as the logistic function, squeezes the input values into the range of [0, 1]. Sigmoid is useful in binary classification tasks, as it can be interpreted as a probability. However, it suffers from the vanishing gradient problem when dealing with deep networks.

- **SoftMax:** Softmax is often used as the activation function in the output layer of a CNN when dealing with multi- class classification problems. It produces a probability distribution over the classes, ensuring that the sum of the probabilities is equal to 1. Softmax converts the output logits into a probability distribution by exponentiating and normalizing them.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Fig. 9 An equation to calculate SoftMax Function

Where,

z represents the values from the neurons of the output layer. The exponential acts as the non-linear function. Later these values are divided by the sum of exponential values in order to normalize and then convert them into probabilities.

The choice of activation function depends on the specific problem, network architecture, and empirical observations. ReLU and its variants, such as Leaky ReLU and PReLU, are widely used due to their simplicity, computational efficiency, and good performance in practice. However, it is worth experimenting with different activation functions to find the one that works best for a given task and network configuration.

Fully Connected Layers: After several convolutional and pooling layers, CNNs often include one or more fully connected layers at the end. These layers connect every neuron in the current layer to every neuron in the subsequent layer, like traditional neural networks. Fully connected layers are responsible for making the final predictions based on the learned features.

The training process of a CNN involves forward propagation, where the input data passes through the network, and the output is compared to the desired output using a loss function (e.g., cross-entropy loss). Then, backpropagation is performed to update the weights of the network based on the error. This iterative process continues until the model learns the underlying patterns and minimizes the loss.

The strength of CNNs lies in their ability to capture spatial relationships in images efficiently. By applying convolutional filters across different regions of an image, CNNs can learn features at different scales and levels of abstraction. This hierarchical feature extraction allows CNNs to recognize complex patterns and objects.

CNNs have been instrumental in numerous computer vision tasks. For instance, in image classification, a CNN can accurately classify images into different categories, such as identifying objects or recognizing handwritten digits. In object detection, CNNs can localize and classify multiple objects within an image. In image segmentation, CNNs can assign a label to each pixel, enabling precise object delineation.

In summary, CNNs are a powerful class of neural networks specifically designed for processing grid-like data, such as images. They leverage convolutional layers to automatically learn hierarchical representations, enabling them to excel in various computer vision tasks. CNNs have significantly advanced the field of computer vision and continue to be an active area of research and innovation.

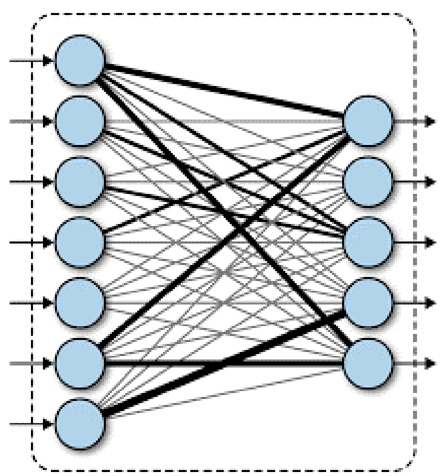


Fig. 9 Dense or Fully Connected layer

The training process of a CNN involves forward propagation, where the input data passes through the network, and the output is compared to the desired output using a loss function (e.g., cross-entropy loss). Then, backpropagation is performed to update the weights of the network based on the error. This iterative process continues until the model learns the underlying patterns and minimizes the loss.

The strength of CNNs lies in their ability to capture spatial relationships in images efficiently. By applying convolutional filters across different regions of an image, CNNs can learn features at different scales and levels of abstraction. This hierarchical feature extraction allows CNNs to recognize complex patterns and objects.

CNNs have been instrumental in numerous computer vision tasks. For instance, in image classification, a CNN can accurately classify images into different categories, such as identifying objects or recognizing handwritten digits. In object detection, CNNs can localize and classify multiple objects within an image. In image segmentation, CNNs can assign a label to each pixel, enabling precise object delineation.

In summary, CNNs are a powerful class of neural networks specifically designed for processing grid-like data, such as images. They leverage convolutional layers to automatically learn hierarchical representations, enabling them to excel in various computer vision tasks. CNNs have significantly advanced the field of computer vision and continue to be an active area of research and innovation.

Final Output Layer: In the convolution layer, neurons are only connected to a local region, whereas in a region with full connectivity, all inputs are connected to neurons. After obtaining values from the fully connected layer, we'll connect them to the final layer of neurons [with a count equal to the total number of classes], which will determine the probability that each image belongs to a distinct class.

3.4 Proposed Model

This deep learning model is initialised as a sequence of layers, so the proposed model is sequential. The model propagates the information from the input layer to the hidden layer to the output layer. This network uses the mathematical operation known as convolution.

The primary purpose of convolution is to locate features in an image using a feature detector and store them in a feature map. It also preserves the spatial relationship between pixels, which is crucial for our purposes. The first step in developing a model for our work is to initialise it. We initialised our model as sequential because it is sequential. The next stage is to construct the first convolutional layer of the CNN.

The input layer of our CNN was created. We have obtained 32 3x3 feature detectors.The layer accepts input for a collection of images. In our case, there are 32 units per order. Our image for the input layer has the dimensions (28,28,1), where the height and width are both 28 and there is one channel because the images are grayscale. After performing a convolution operation on the image, each feature detector generates a feature map with an output volume of 26x26x1.

We've used the default values for padding (0) and stride (1). After convolution, the ReLU activation function is applied to our CNN model to increase non-linearity. If the value is greater than 0, the input is accepted; otherwise, the function returns 0. Now that Batch Normalisation has been conducted to maintain the layer's output within the same range, we have an unbiased result. The output of the first convolution layer is then transmitted to the next layer for the max-pooling operation. Max pooling helps eliminate superfluous features and accounts for their spatial, textual, and other distortions. Pooling also prevents overfitting by omitting a few parameters when generating a pooled matrix during the pooling operation.

Max pooling is utilised to isolate only the pertinent image information by extracting only higher values from a portion of the input using an empty feature detector. Down sampling is another name for pooling. We take stride as two and the dimension of the filter feature detector as 2x2. Consequently, the height and breadth are diminished by a factor of 2.

The second layer consists of a convolution 2d layer and a max pooling layer. This layer receives the output of the preceding layer as its input. The nucleus size is 3x3 and there are 64 unique feature detectors. We then reapplied Batch Normalisation and passed the result as an input to Max Pooling. Now we have the last layer, which is identical to all previous layers in that it consists of 1 conv2d layer and 1 max pooling layer, so the operations conducted will be similar; however, there are 128 feature detectors in the last layer.

Now, the output of this dense or hidden layer is fed to the 25-neuron dense output layer. This layer contains 25 neurons that correspond to 25 classes. We have utilised the Softmax activation function in this instance. As the sigmoid function is the core of the probabilistic approach, we could have used it here. However, the sigmoid function is only useful when classifying between two classes. Therefore, when there are more classes, SoftMax is used, and SoftMax ensures that the sum of the probabilities of the output is one, making it simple to comprehend.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Fig. 4 An equation to calculate SoftMax Function

Where,

z represents the values of the output layer neurons. Exponentiation serves as the nonlinear function. Later, these values are normalised by dividing them by the sum of exponential values, and then they are converted into probabilities

SoftMax transforms the linear input data into a probability array of all 25 classes, checks the probabilities against the actual output, and calculates the loss utilising sparse categorical cross-entropy. We have employed sparse categorical cross-entropy in this instance because we have labelled as targets.

Finally, we compile our model and use the 'Adam' optimizer (stochastic Gradient Descent Method) to achieve a global minimum in order to minimise our cost function. The 'Adam' optimizer is used with a 0.001 learning rate and b1=0.9 and b2=0.99 values.

CHAPTER 4

RESULTS AND DISCUSSION

MNIST ASL data is used to train the CNN model. The training dataset consists of 27455 training samples with 784 features. Utilising cross-entropy ADAM, the model was trained to minimise loss. The varied model is trained for 50 iterations/epochs with 512 batches. The model was trained at a rate of 0.01 . There are 7172 samples in the validation dataset, and the validation accuracy of the model reportedly exceeds 97%. The following are the accuracy and loss plots for training.

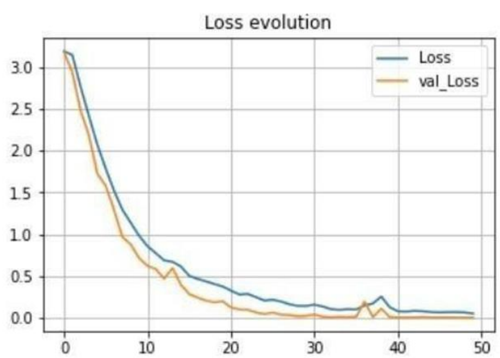


Fig. 5 Graph plots of loss evolution

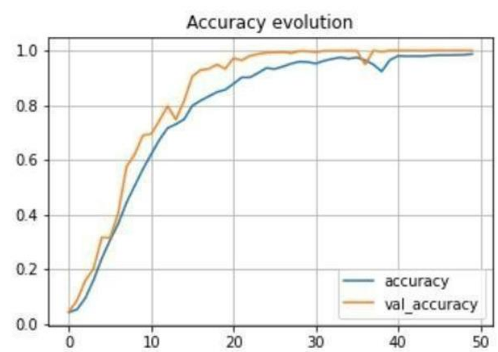


Fig. 6 Graph plot of accuracy evolution

In the loss evolution graph, the loss value is decreasing with time and in accuracy evolution plot, the accuracy is increasing with time which means that our model is learning.

CONFUSION MATRIX:

In machine learning and statistics, a confusion matrix is a performance evaluation instrument used to assess the performance of a classification model. It provides a tabular representation of a model's predictions against the actual class labels of a dataset.

Typically, the confusion matrix is a square matrix that illustrates the four essential metrics:

1. True Positives (TP): The number of instances accurately predicted by the model as positive (class label).

2. True Negatives (TN): The number of instances accurately predicted by the model as negative (not the class label).

3. False Positives (FP): The number of instances that the model incorrectly classifies as positive (class label). Also referred to as a Type I error.

4. False Negatives (FN): The number of instances that were incorrectly predicted by the model as negative (not the class label). Also referred to as a Type II defect.

By categorising the predictions into these four categories, the confusion matrix provides a comprehensive overview of the model's performance. It assists in comprehending the varieties of model errors, such as false positives and false negatives. This data is essential for evaluating the model's precision, recall, and other performance metrics.

The graph generated during training, along with the corresponding matrices for Model Accuracy and Confusion, is depicted in Figure 7.

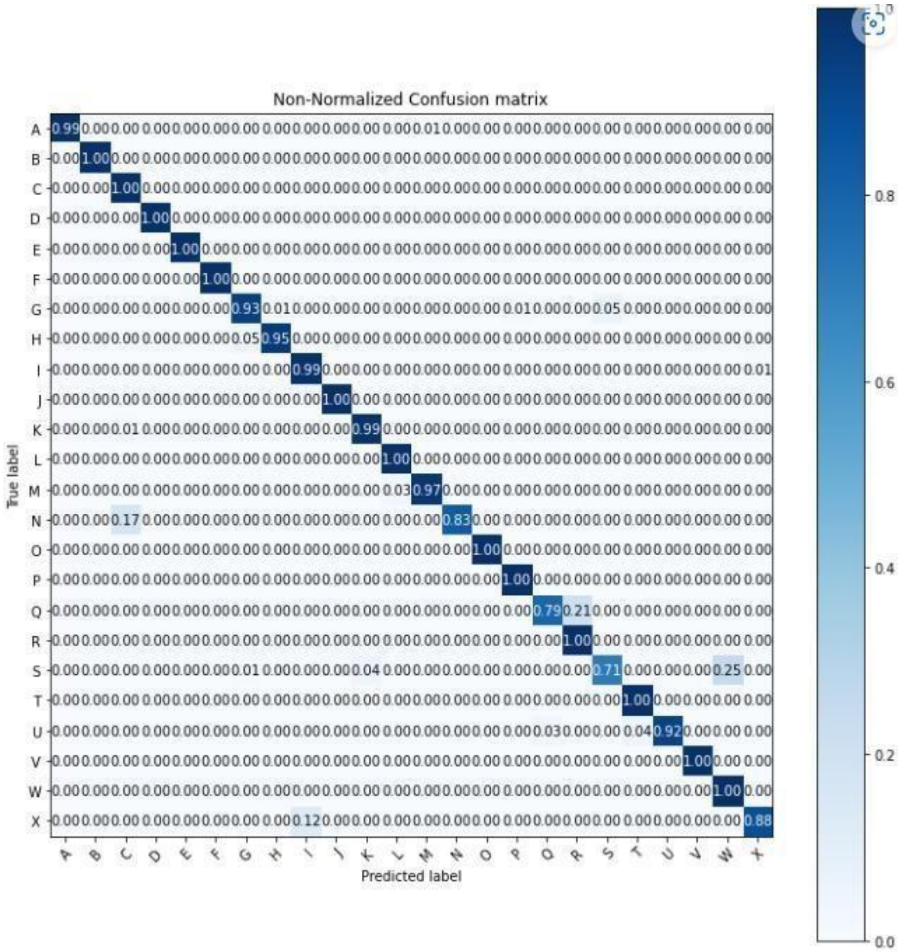


Fig. 11 Confusion Matrix

The CNN model has given 98% accuracy in class label prediction for 12 classes, as we can see in the above figure.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

This paper presents a CNN-based method for recognising and classifying sign language using computer vision. Unlike other approaches, this method yields higher precision than [10].

5.2 FUTURE SCOPE

Future scope includes, but is not restricted to:

- This paradigm can be applied to other sign languages, such as Indian Sign Language, but for the time being it is exclusive to American Sign Language.
- The model can be trained with an additional data set to autonomously segment the gesture from the captured frame by subtracting the background automatically.
- Tuning and enhancing the model to recognise common words and expressions
- Additionally, training the neural network model to identify symbols efficiently requires two hands.
- Include active hand gestures as a supplement to the current static finger lettering.
- Integration of the optimised model into existing AI systems, such as Amazon Alexa, for visual recognition advancements

REFERENCES

[1] Ss, Shivashankara & S, Dr.Srinath. (2018). American Sign Language Recognition System: An Optimal Approach. International Journal of Image, Graphics and Signal Processing. 10. 10.5815/ijigsp.2018.08.03.

[2] Rokade, Yogeshwar & Jadav, Prashant. (2017). Indian Sign Language Recognition System. International Journal of Engineering and Technology. 9. 189-196. 10.21817/ijet/2017/v9i3/170903S030.

[3] Sutton-Spence, Rachel & Woll, Bencie. (2008). British Sign Language. 10.1002/9780470757000.ch6.

[4] https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[5] S. Shahriar et al., "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning," TENCON 2018 - 2018 IEEE Region 10 Conference, 2018, pp. 1168-1171, doi: 10.1109/TENCON.2018.8650524.

[6] Jain, Vanita & Jain, Achin & Chauhan, Abhinav & Kotla, Srinivasu & Gautam, Ashish. (2021). American Sign Language recognition using Support Vector Machine and Convolutional Neural Network. International Journal of Information Technology. 13. 10.1007/s41870-021- 00617-x.

[7] Kaggle (2017) Sign language MNIST: drop-in replacement for MNIST for hand gesture recognition tasks. https://www.kaggle. com/datamunge/sign-language-mnist. Accessed Jan 2018

[8] Rasha Amer Kadhim 1 , Muntadher Khamees . “A Real-Time American Sign Language Recognition System using Convolutional Neural Network for Real Datasets”.doi: 10.18421/TEM93-14, August 2020

[9] Real-time American Sign Language Recognition with Convolutional Neural Networks . Brandon Garcia Stanford University Stanford, CA. Sigberto Alarcon Viesca Stanford University Stanford, CA.

[10] R.S, Dr.Sabeenian & Bharathwaj, S. & Aadhil, M.. (2020). Sign Language Recognition Using Deep Learning and Computer Vision. Journalof Advanced Research in Dynamical and Control Systems. 12. 964-968. 10.5373/JARDCS/V12SP5/20201842.

[11] Kasukurthi, Nikhil, Brij Rokad, Shiv Bidani, and Dr Dennisan. "American Sign Language Alphabet Recognition using Deep Learning." arXiv preprint arXiv:1905.05487 (2019).

[12] Zimmermann, C., & Brox, T. (2017). Learning to Estimate 3D Hand Pose from Single RGB Images. arXiv preprint arXiv:1705.01389.

[13] Rahman, Md Moklesur, Md Shafiqul Islam, Md Hafizur Rahman, Roberto Sassi, Massimo W. Rivolta, and Md Aktaruzzaman. "A new benchmark on american sign language recognition using convolutional neural network." In 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), pp. 1-6. IEEE, 2019.

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

[15] F. Bes¸er, M. A. Kizrak, B. Bolat, and T. Yildirim, “Recognition of sign language using capsule networks,” in 2018 26th Signal Processingand Communications Applications Conference (SIU). IEEE, 2018, pp. 1–4.

[16] B. Garcia and S. A. Viesca, “Real-time american sign language recognition with convolutional neural networks,” Convolutional NeuralNetworks for Visual Recognition, vol. 2, 2016.

[17] A. Deza and D. Hasan, “Mie324 final report: Sign language recognition.”

[18] ” A Real-Time System For Recognition Of American Sign Language By Using Deep Learning” Murat Taskiran, Mehmet Killioglu, and Nihan Kahraman Electronics and Communications Engineering Yildiz Technical University Istanbul, Turkey

[19] Chavan, Shruti, Xinrui Yu, and Jafar Saniie. "Convolutional Neural Network Hand Gesture Recognition for American Sign Language." In 2021 IEEE International Conference on Electro Information Technology (EIT), pp. 188-192. IEEE, 2021.

[20] Rao, G. Anantha, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry. "Deep convolutional neural networks for sign language recognition." In 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), pp. 194-197. IEEE, 2018.

[21] Sharma, Sakshi, and Sukhwinder Singh. "Vision-based hand gesture recognition using deep learning for the interpretation of sign language." Expert Systems with Applications 182 (2021): 115657.

[22] Carneiro, AL Cavalcante, L. Brito Silva, and DH Pinheiro Salvadeo. "Efficient sign language recognition system and dataset creation method based on deep learning and image processing." In Thirteenth International Conference on Digital Image Processing (ICDIP 2021), vol. 11878, pp. 11-19. SPIE, 2021.

[23] Jie Huang, Wengang Zhou, Houqiang Li and Weiping Li, "Sign Language Recognition using 3D

convolutional neural networks," 2015 IEEE International Conference on Multimedia and Expo (ICME), 2015, pp. 1-6, doi: 10.1109/ICME.2015.7177428.

[24] Varma, Teena, Ricketa Baptista, Daksha Chithirai Pandi and Ryland Coutinho. "Sign Language Detection using Image Processing and Deep Learning." (2020).

[25] Pigou, Lionel, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. "Sign language recognition using convolutional neural networks." In European conference on computer vision, pp. 572-578. Springer, Cham, 2014.

APPENDIX

Deep Learning Technique for Detection of Sign Language

[1]Prof. Bharti, [2]Himanshi Tyagi, [3]Shreya Shankar, [4]Surbhi Sawan

[1]Assistant Professor, KIET Group of Institutions, Ghaziabad, U.P.
[2]Student, KIET Group of Institutions, Ghaziabad, U.P.
[3]Student, KIET Group of Institutions, Ghaziabad, U.P.
[4]Student, KIET Group of Institutions, Ghaziabad, U.P.

Email: [1]bharti.cse@kiet.edu, [2]himanshi.1923cs1018@kiet.edu, [3]shreya.1923cs1010@kiet.edu, [4]surbhi.1923cs1015@kiet.edu
Contact: [1]7206459574, [2]7668921558, [3]8851768230, [4]7668551616

***Abstract*—** Language is a medium through which people can communicate and share their thoughts and words with each other and Sign language is one of the languages in which people make use of hand movements and gestures to express themselves. Normal people have trouble understanding and interpreting sign language's meaning. It really has become necessary to understand the sign language, so we need an interpreter, as expressed by the hearing impaired. The main facilitator in assisting hard-of-hearing people in interacting with the rest of society is learning sign language. The process by which a computer analyses and converts sign language gestures into understandable and human-readable text is known as "sign language detection." Those who have trouble hearing or speaking can communicate with ease by using Sign language detection software. Many different sorts of studies are being done to make this process easy and efficient. In this paper, we try to highlight the work and comparative study of that work done by researchers in American Sign Language using deep learning.

***Keywords*—** American Sign Language (ASL), Deep Learning (DL), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Principle Components Analysis (PCA), and Histogram Of Oriented Gradients (HOG).

1. Introduction

To communicate words and ideas using gestures and hand movement instead of spoken words is sign language. It is used by deaf people and others who want to communicate using only their hands. Members of the sign language community use signs as a means of communicating with each other and with others who are not deaf. Sign language is also used as a way to express oneself in public; for instance, when a person wants to express thanks, apologize, or protest, they may make signs using their hands. There are numerous sign languages across the globe, including American Sign Language [1] (ASL) in America, Indian Sign Language[2] (ISL) in India, and German Sign Language (GSL) in Germany. But in this paper, we have used ASL instead of any other sign language.

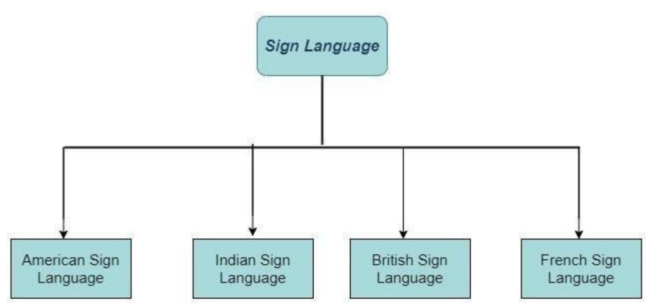


Fig.1 Types of sign language.

Table.1 Comparison Between Various Sign Languages

Sign Language	Comparison		
	Similarities	***Differences***	***Region***
American Sign Language [1]	Involve static and dynamic gestures	Use only the one-handed manual alphabet in ASL.	USA and parts of Canada.

Sign Language	Comparison		
	Similarities	*Differences*	*Region*
		Use only fingers to represent vowels.	
Indian Sign Language [2]	Involve static and dynamic gesture	Use both hand to show letters.	South Asian Countries.
British Sign Language [3]	almost similar to ISL.(two-handed used manually alphabet in BSL).	use thumb and four finger to represent vowels.	England & Northern Ireland.

American Sign Language [1] is the most widely used sign language. It is mostly spoken in the United States and parts of Canada, among many other countries. Most researchers prefer it because it can be easily classified and recognized due to the one-handed representation of the alphabet in ASL. While in Indian Sign Language[2] and British Sign Language[3], we used a two-handed manual alphabet, which makes it difficult to classify and recognize letters in ISL and BSL. So, ASL is the most preferred language all over the world.

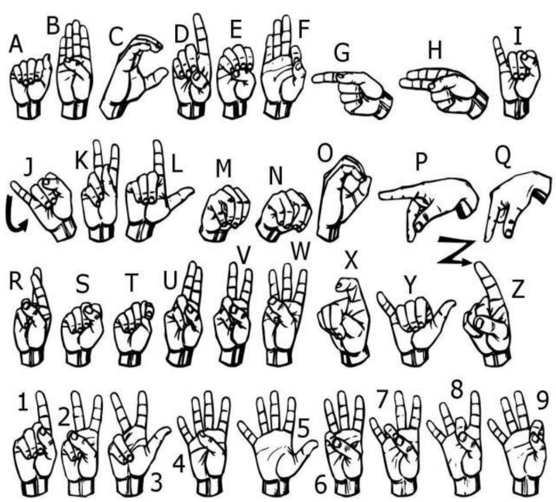


Fig. 2. Basic Gestures in American Sign Language [20]

2. Literature Review

According to the WHO [4], more than 5% of the world's population requires therapy to heal their "disabling" hearing loss (432 million adults and 34 million children). It is roughly calculated that by 2050, over 700 million people, or one in every ten people, will have a disabling hearing loss. For this reason, studies are done to help disabled people to communicate more easily.

In Ref[5] Real-Time American SLR Using Skin Segmentation and Image Category Classification with CNN and DL, a hand-signs image dataset is used where CNN and deep learning techniques are used, and the overall test accuracy of the paper is 94.7%. In Ref[6], American Sign Language recognition using SVM and CNN, in which the MNIST Kaggle dataset is used, SVM with a "poly" kernel and double-layer CNN techniques are used, and the accuracy of the paper is 81.49% in SVM and 98.58% in CNN. In Ref[8], A Real-Time ASL Recognition System Using CNN for real datasets, the Hand Gesture Images dataset is used where the CNN technique is applied, and the maximum accuracy of the paper is 98.53%. In Ref[9], the ASL Finger-spelling Dataset from the University of Surrey's Centre for Vision is used in Real-time American SLR with CNN, and the accuracy of the paper is 98% with five letters and 74% with ten letters. The MNIST-Kaggle dataset [7] is used to implement SLR Using DL and Computer Vision in Ref[10], and the accuracy rate of the paper is more than 93%. The Squeeze Net Model approach is employed in Ref[11] ASL Alphabet Recognition using DL, which uses the Surrey Finger dataset [12], and yields an accuracy rate of 83.29%. The Massey University Gesture dataset [14], the sign language digit dataset [15], the ASL finger spelling dataset [16], and the ASL alphabet dataset [17] are used in Ref[13], a new benchmark on American SLR using CNN, where the CNN technique is applied and 100% precision is attained for the MU Hand Image ASL dataset. The Institute of Information and Mathematical Sciences acquired the dataset for A Real-Time System For Recognition Of ASL utilizing DL in 2011, and CNN approach was utilized to obtain 98.05% test performance [14]. The ASL dataset [15] was utilized in Ref[19], which used CNN Hand Gesture Recognition for ASL, and test accuracy of 87.5% was accomplished. In Ref[20], Deep CNN for SLR model is executed on Own Hand-Sign Image Dataset with 92.88% recognition rate. In Ref[21], vision-based hand gesture recognition using DL for sign language assessment is used, and indeed the G-CNN, VGG-11, as well as VGG-16techniquesare used, with the G-CNN model obtaining the highest classification performance of 94.83%, 99.96%, and 100%, respectively. In Ref[22], the own hand-sign video dataset is used in an efficient SLR system and dataset creation method based on DL and image processing, where the multi-stream CNN strategy is being used, with precision of 96%onthetestset and 81% on the validation data. In Ref[23], the average accuracy in SLR using 3D CNN, where the own sign language video dataset was used and the 3D CNN technique was used, was 94.2%. In Ref[24], Sign Language Detection using Image Processing and DL, where an ASL image dataset is used and the CNN technique is applied, the Canny Edge algorithm with an accuracy of 98% demonstrated better results than the other techniques. In Ref[25], SLR Using CNN, where a dataset from ChaLearn Looking at People 2014 [26] is used and CNN and GPU acceleration techniques are applied, a validation accuracy of 91.70% is obtained.

3. Preliminaries

3.1 Data Pre-processing

In the paper, ASL is used, and the dataset is fetched from MNIST, which is published on Kaggle [7]. It is composed of training images and test images, which are 27455 and 7172, respectively all with 28x28 pixel geometry. These images belong to 25 classes from A to Y in the English alphabet (no Z class designation due to gesture movement). The Kaggle dataset is in CSV format, where rows and columns are 27455 and 785, respectively, of training data. The image class is represented by the first column, and the rest all represent 28x28 pixels. The test data set follows the same paradigm. Before the input data can be fed into the model, it needs to be preprocessed. Pre-processing is performed to improve data quality and speed up training. The image was transformed to a grayscale image. The highest value of the gray level range is used to normalize the gray levels in the input image. After this transformation, data splitting was performed witharatioof80:20. CNN accepts input in a specific format, so the train date was changed to (21964, 28, 28, 1).

3.2 Deep Learning Approach

ConvNet is another name of the CNN. It is an ANN that has so far been employed for image processing and object detection. It can also be applied for classification tasks or data analysis. It generally has some kind of specialization so that it can select out or discover patterns that make sense of it. CNN's ability to identify patterns is what makes it so beneficial for image analysis.

Pixels are used as a value in the input layer of a matrix that contains images as input. Each pixel will have a value between 0 and 255. The "hidden layer" is another name for the convolutional layer. It detects patterns, edges, shapes, corners, and various objects using filters. There are certain kinds of operations such as convolution, padding, stride, and filter (in kernels). In convolutional operations, a convolution converts all the pixels in its receptive field into a single value. The filter is placed on top of the input matrix, then we perform dot product, and finally we produce the output matrix by performing the operation. The number of rows and columns of a matrix with values that are initialized with random numbers is chosen, even though a filter can technically be regarded of as nothing more than a small matrix. Actually, it extracts the information from the specific image. While padding is about building the compound around the images, it is like applying a layer on top of the image. The main purpose of this operation is to prevent the loss of information due to convolution. In the Stride operation, we basically compress images and video data. It is the quantity of pixels that move across the input matrix. Equation represents the output matrix following convolution

$$output = \frac{(N+(2*P)-F+1)}{S}$$

where N, P, F, S stands for input matrix size, padding, filter matrix size, and stride, respectively. In pooling, layers are employed to minimize the convolution output's dimensionality. There are three types of pooling: maximum, minimum, and average. Consequently, it decreases the quantity of computation done in the network as well as the amount of parameters to learn. The components of a fully connected layer include neurons, weights, and biases. It links the neurons in one layer with those in another layer, allowing the data to be classified into different categories by flattening and dropout training. The required prediction is attained at the output layer, which is the last layer. It adds its weights and biases before determining the final output.

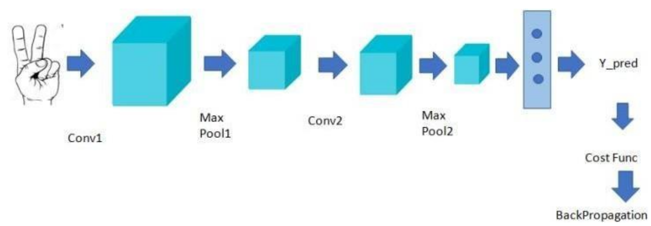


Fig.3 Architecture of a CNN Convolution Operation

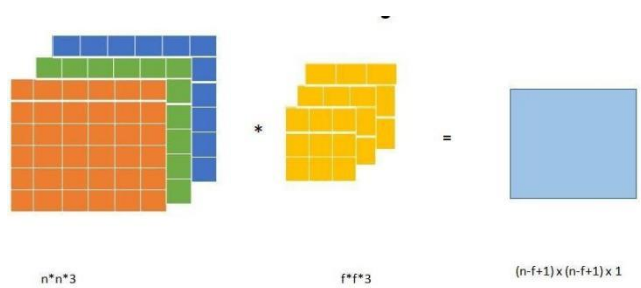


Fig.4 Convolution Operation

3.3 Performance Evaluation Metrics

Loss Plot: The loss function is also known as the cost function or error function, calculating the variance between the desired value and algorithm output is its primary purpose. If a high error rate is thrown, a heavy loss will occur, which means that the model did not perform well. Otherwise, the lower the error rate, the better our model works. The loss function is utilized by the model in deep learning to acquire new information. The objective is to minimize loss, which can be done using gradient descent, which modifies the model's parameters based on the loss outcome.

(

$$Loss = -\sum_{j=1}^{k} y_j \log\left(\hat{y}_j\right)$$

where K stands for the number of data classes in the data, y for actual value, and z for neural network prediction.

Accuracy Plot: Accuracy describes how the model is going to perform in all classes. It compares model predictions to true values using a percentage to determine how well our model predicts. With low accuracy and high loss, the model produces big errors, and the model will have minor errors if accuracy and loss both are low. However, if both are high, it will cause big errors. Lastly, the model will make minor errors ononlysome of the data if accuracy is intense and loss is low.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Where, TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.

4. Proposed Methodology

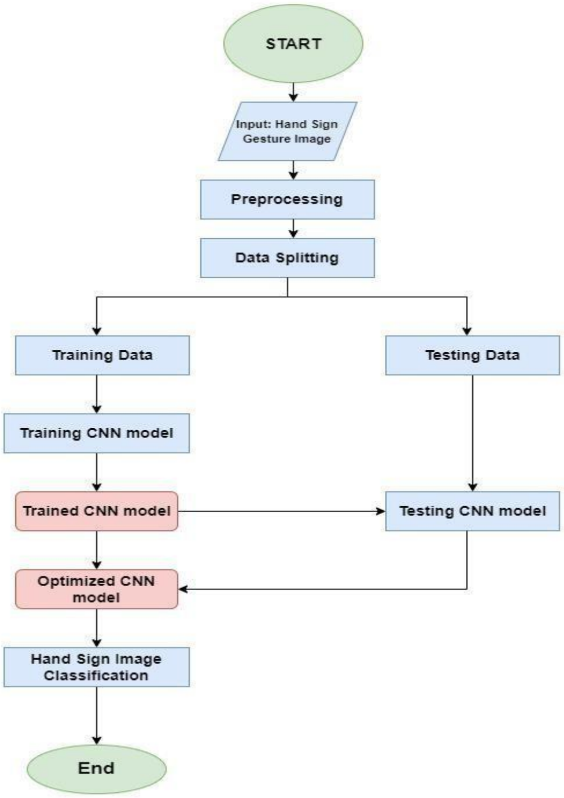


Fig.5 Flowchart of Proposed Model

In the above flow chart, we have shown our proposed methodology for sign language recognition.We preprocessed the dataset and then split it into training and testing dataset. Then we train our model with training dataset. Thereafter, we test our trained CNN model with testing dataset and optimized it.

5. Results & Discussion

The deep learning model propagates data through a series of layers from the layer of input to the hidden layer, and finally to the output layer, information is transmitted through a sequence of layers in the DL model. The mathematical operation that is used is convolution. Its primary goal is to determine the features in an image using a feature detector and to preserve the relationship between pixels in an image. Initially, a working model was created to initialize it. Initialization is done sequentially as the model is sequential. Secondly, the CNN layer is created to perform the convolution operation.

As the input layer is created for the CNN, it obtains 32 feature detectors of size 3x3. Then input is provided to the layer, which converts the input into batches of images with 32 sizes. The image has the form of (28, 28), where height is 28 and width is 28. A channel is present because of the gray scale image. A feature map is produced by a feature detector after convolution operation on an image with an output volume of 26x26x1. By default, padding is set to zero and stride is set to one. Following that, the ReLu function is executed to increase the non-linearity of the model. If value is positive, the input is passed else, return zero. The batch is normalized to produce output in the same range as the unbiased result. The result of the first layer of convolution is transmitted to the following layer for maximum pooling. It will help eliminate unwanted features and noise. It avoids those parameters that cause overfitting. Max pooling extracts useful information from images by using an empty feature detector and producing higher values from it. Set stride to 2 and feature detectorto2x2, and then height and width are reduced by the 2 factors. A max pooling layer and a 2D convolutional layer are present in the second layer. The previous layer's output serves as the following layer's input. 3x3 size kernel, and 64differentfeature detectors are present. Batch normalizations were executed again and passed to the max pooling layer. At the final level, similar operations are performed on the same convolutional layer and max pooling layer, but they differ by the feature detector, which is 128. A flattening operation is performed to eliminate the negativity using the Relu. Utilizing the Relu function, another 10 dense layers are executed to increase the nonlinearity. Then 25% of random neurons are dropped out by the dropout layer to prevent overfitting. The final layer of output receives input from the concealed layer, which is composed of 25 neurons. These 25 neurons represent 25classes. Here, the Softmax Activation function is used. When there are more than two classes, the Softmax function is used. It is easy to understand if the output probabilities sum to one.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

Equation of SoftMax Function

.

The SoftMax function alters the linear data into an array of 25 classes. For the targeted sparse category cross-entropy, the actual output and loss probability are computed. Finally, on a compiled model, an Adam optimizer is used to reach the minimum cost function. The CNN model is trained against the MNIST dataset, where 27455 are training samples and 784 train features. Cross-entropy ADAM is used for the model to reduce loss. 512 batches of 50 epochs each are used to train the model, which has a 0.001 learning rate. The accuracy of the validation dataset, which has 7172 samples, is 97%.

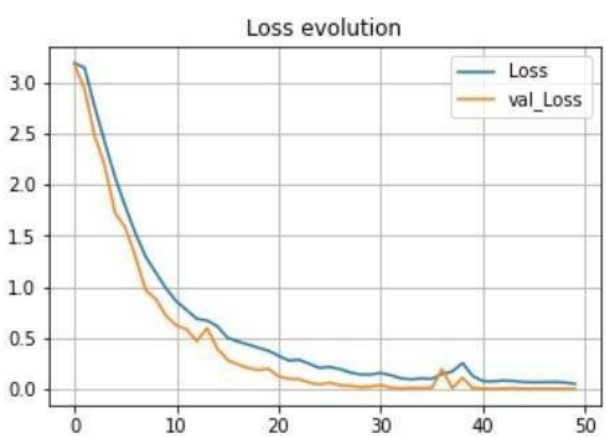


Fig. 5 Graph plots of loss evolution

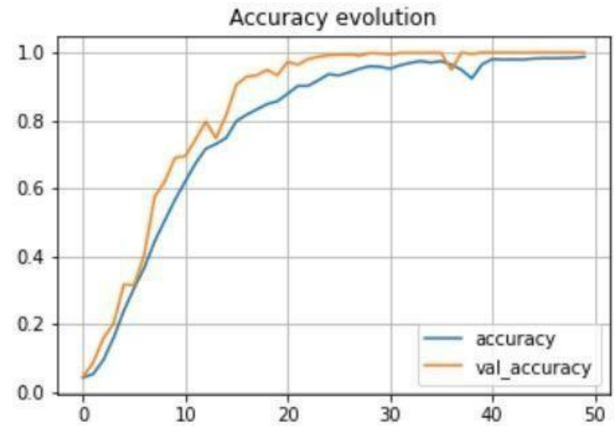


Fig. 6 Graph plots accuracy evolution

In the loss evolution graph, the loss value is decreasing with time and in accuracy evolution plot, the accuracy is increasing with time which means that our model is learning.

6. CONCLUSION

In this research, a DL method for the recognition and categorization of sign language that is CNN-based is provided. This method's accuracy is more than that of studies by R.S. Sabeenian, S. Sai Bharathwaj, and M. Mohamed Aadhil published in their paper SLR using DL and Computer Vision [10]. This model can be trained further such that the background is removed, and the hand motion is instantly separated from the captured frame. Due to their dynamic hand motions, "J" and "Z" letters were excluded from th classification. They require video frames for classification. Further research can be done to capture video frames and classify the dynamic letters "J" and "Z."

REFERENCES

[1] Ss, Shivashankara & S, Dr.Srinath. (2018). American Sign Language Recognition System: An Optimal Approach. International Journal of Image, Graphics and Signal Processing. 10. 10.5815/ijigsp.2018.08.03.

[2] Rokade, Yogeshwar & Jadav, Prashant. (2017). Indian Sign Language Recognition System. International Journal of Engineering and Technology. 9. 189-196. 10.21817/ijet/2017/v9i3/170903S030.

[3] Sutton-Spence, Rachel & Woll, Bencie. (2008). British Sign Language. 10.1002/9780470757000.ch6.

[4] https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[5] S. Shahriar et al., "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning," TENCON 2018 - 2018 IEEE Region 10 Conference, 2018, pp. 1168-1171, doi: 10.1109/TENCON.2018.8650524.

[6] Jain, Vanita & Jain, Achin & Chauhan, Abhinav & Kotla, Srinivasu & Gautam, Ashish. (2021). American Sign Language recognition using Support Vector Machine and Convolutional Neural Network. International Journal of Information Technology. 13. 10.1007/s41870-021-00617-x.

[7] Kaggle (2017) Sign language MNIST: drop-in replacement for MNIST for hand gesture recognition tasks. https://www.kaggle.com/datamunge/sign-language-mnist. Accessed Jan 2018

[8] Rasha Amer Kadhim 1 , Muntadher Khamees . "A Real-Time American Sign Language Recognition System using Convolutional Neural Network for Real Datasets".doi: 10.18421/TEM93-14, August 2020

[9] Real-time American Sign Language Recognition with Convolutional Neural Networks . Brandon Garcia Stanford University Stanford, CA. Sigberto Alarcon Viesca Stanford University Stanford, CA.

[10] R.S, Dr.Sabeenian & Bharathwaj, S. & Aadhil, M.. (2020). Sign Language Recognition Using Deep Learning and Computer Vision. Journal of Advanced Research in Dynamical and Control Systems. 12. 964-968. 10.5373/JARDCS/V12SP5/20201842.

[11] Kasukurthi, Nikhil, Brij Rokad, Shiv Bidani, and Dr Dennisan. "American Sign Language Alphabet Recognition using Deep Learning." arXiv preprint arXiv:1905.05487 (2019).

[12] Zimmermann, C., & Brox, T. (2017). Learning to Estimate 3D Hand Pose from Single RGB Images. arXiv preprint arXiv:1705.01389.

[13] Rahman, Md Moklesur, Md Shafiqul Islam, Md Hafizur Rahman, Roberto Sassi, Massimo W. Rivolta, and Md Aktaruzzaman. "A new benchmark on american sign language recognition using convolutional neural network." In 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), pp. 1-6. IEEE, 2019.

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

[15] F. Bes¸er, M. A. Kizrak, B. Bolat, and T. Yildirim, "Recognition of sign language using capsule networks," in 2018 26th Signal Processing and Communications Applications Conference (SIU). IEEE, 2018, pp. 1–4.

[16] B. Garcia and S. A. Viesca, "Real-time american sign language recognition with convolutional neural networks," Convolutional Neural Networks for Visual Recognition, vol. 2, 2016.

[17] A. Deza and D. Hasan, "Mie324 final report: Sign language vrecognition."

[18] " A Real-Time System For Recognition Of American Sign Language By Using Deep Learning" Murat Taskiran, Mehmet Killioglu, and Nihan Kahraman Electronics and Communications Engineering Yildiz Technical University Istanbul, Turkey

[19] Chavan, Shruti, Xinrui Yu, and Jafar Saniie. "Convolutional Neural Network Hand Gesture Recognition for American Sign Language." In 2021 IEEE International Conference on Electro Information Technology (EIT), pp. 188-192. IEEE, 2021.

[20] Rao, G. Anantha, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry. "Deep convolutional neural networks for sign language recognition." In 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), pp. 194-197. IEEE, 2018.

[21] Sharma, Sakshi, and Sukhwinder Singh. "Vision-based hand gesture recognition using deep learning for the interpretation of sign language." Expert Systems with Applications 182 (2021): 115657.

[22] Carneiro, AL Cavalcante, L. Brito Silva, and DH Pinheiro Salvadeo. "Efficient sign language recognition system and dataset creation method based on deep learning and image processing." In Thirteenth International Conference on Digital Image Processing (ICDIP 2021), vol. 11878, pp. 11-19. SPIE, 2021.

[23] Jie Huang, Wengang Zhou, Houqiang Li and Weiping Li, "Sign Language Recognition using 3D convolutional neural networks," 2015 IEEE International Conference on Multimedia and Expo (ICME), 2015, pp. 1-6, doi: 10.1109/ICME.2015.7177428.

[24] Varma, Teena, Ricketa Baptista, Daksha Chithirai Pandi and Ryland Coutinho. "Sign Language Detection using Image Processing and Deep Learning." (2020).

[25] Pigou, Lionel, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. "Sign language recognition using convolutional neural networks." In European conference on computer vision, pp. 572-578. Springer, Cham, 2014.