IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Survey On Lip Reading Recognition Using Artificial Intelligence

Shubham Kalburgi[1], Prof. Vina M. Lomte[2], Niraj Jain[3], Ashutosh Buge[4], Shreya Kankal[5]

[1-5]Department of Computer Engineering, RMD Sinhgad School of Engineering, SPPU, India.

Abstract – Lip Reading Recognition System is a technology that uses Artificial Intelligence to recognize words or speech by visual interpretation of face, mouth and lip movement without involvement of audio. Imagine you are talking to someone, and they can't hear what you're saying. They might try to understand your words by watching how your lips move. That's called lip reading. This task is difficult as people use different dictions and various ways to articulate a speech. The system works by using a camera to capture a video of a person's face while they are speaking. It's like a recording a video of their lips moving. The video is then analyzed by an AI algorithm that can identify lip movements, extract features, train the classifier by giving user's lip movement frames sequences as input and will identify said word and finally translate them into words. It's not always perfect, but it's often quite accurate. This can help deaf and mute individuals in understanding spoken language through visual cues of the speaker's lips and addressing communication barriers in their daily lives. It can also be used to improve privacy and security in applications for detecting suspicious behavior. The survey focuses on deep learning related approaches such as convolutional neural networks (CNN) [6], recurrent neural networks (RNN) [15] and their variants which have been proven to be more fruitful for both feature extraction and classification. This paper provides comparisons of different audio-visual databases, various algorithmic techniques by highlighting their accuracies, feature extraction methodologies and classification networks.

1. INTRODUCTION

Lip reading is when we understand what people are saying by looking at their lip's moments. Lip reading is an essential skill for some people and has many real-world applications. Lip reading can help people in hearing impairment and have many securities and surveillance uses as well. It can help doctor talk to patients, especially when they can't speak. In school, it can make learning easier. Lip Reading can help in emotion recognition and sentiment analysis and has other applications in different domains. Many Researchers have found the way to implement the Artificial Intelligence and Deep learning techniques for lip reading. There are many techniques developed for image processing in the field for deep Learning and Artificial Intelligence.

In this survey paper we are going to explore lip reading and how the techniques of Artificial Intelligence and Deep Learning helps us to recognize the text just by reading lip movements. We will explore the evolution of the technologies, methods, dataset used in Lip reading. This survey paper aims to provide the comprehensive review on lip reading techniques. We Will take the closer look at how Deep Learning makes sense of Lip movement and turns them into words. Even with technology, lip reading is difficult. The computer becomes confused when there is background noise or poor illumination. Furthermore, a few words have very similar pronunciations, which makes it difficult for the computer to interpret speech. We will examine the transition of lip reading from a specialized field to a widely used technology, the challenges that still need to be addressed, and the fascinating prospects that this field offers through this study. We cordially invite you to join us on this adventure as we explore the world of AI lip reading and its revolutionary possibilities in the field of human-computer interaction and beyond.

This paper also consists of survey of different datasets available on Lip reading as it is very important to test the model on different datasets to get accurate results. As it is observed in most of the model uses combination of 3D-CNN and LSTM has provided more accurate results[1]. For building this model we need to review the different techniques and datasets so that identify how to build more accurate model. This survey reviews the different Artificial Intelligence and Deep Learning Approaches and different datasets as well. So, continue reading to find out how Artificial Intelligence and Deep Learning is affecting lip reading and how it can help both people and technology. This survey will give the comprehensive understanding of lip-reading invention from this survey.

The post is organized as follows - In part 2, we analyzed lip reading challenges. In part 3, we presented a literature review of similar work. In Section 4, we did a detailed survey of the various algorithms. In Section 5, the characteristics of various lip-reading datasets are generally used. We also trace some deep learning network models used to implement the lip-reading system.

2. CHALLENGES IN LIP READING

1. Variability in speech and lip movements: The variety in how people speak and move their lips introduces significant variability. Deep learning models must account for this wide variety of speech patterns, accents, and lip movements, making training and recognition more challenging.

2. Phoneme and Visemes Ambiguity: Many phonemes and visemes (visual counterparts of phonemes) have similar lip movements, leading to ambiguity in recognition. Developing models that can differentiate similar lip patterns is a significant challenge.

3. Generalization to new languages: Extending reading models to new languages and dialects requires additional efforts in data set collection, model fitting, and linguistic knowledge.

4. Overcoming noise and environmental factors: Lip reading in real-world scenarios often faces challenges such as noisy environments, different lighting conditions, and different viewing angles. In order to reliably recognize deep learning models, they must be robust to these environmental factors.

5. Real-time processing: Lip reading for applications such as real-time communication or autonomous vehicles requires low-latency processing. Deep learning models must perform quickly without compromising accuracy.

3. Literature Survey

Table -1: Deep Literature Survey of Current Technologies

Ref.no	Paper Title and Paper publication	Methodology Used	Dataset Used	Accuracy	Research gap Identified / Future Scope
[1] Kuldeep Vayadande, Tejas Adsare, Neeraj Agrawal, Tejas Dharmik, Aishwarya Patil, Sakshi Zod	Title: Lipread Net: A Deep Learning Approach to Lip Reading Journal: IEEE Access	3DCNN (3D Convolutional Neural Network), Bidirectional LSTM (Long Short-Term Memory) to transcribe speech from lip movements.	GRID corpus consists of thousands video clips of thirty-four speakers, each saying 1000 sentences, resulting in a sum of 34,000 sentences.	93%	Future prospects include integrating lipreading into hearing aids or cochlear implants to enhance speech recognition in challenging settings.

[2] Brais Martinez, Pingchuan Ma, Stavros Petridis, Maja Pantic	Title: Lip Reading Using Tempopral Convolutional Networks. Journal: IEEE Access	The Temporal Convolutional Network model is used and also simplify the training procesure for lip reading.	LRW, LRW1000	85.3%	To train the large datasets like LRW and LRW 1000 the model must be strong and the high-quality algorithms to extraction and training of those datasets.
[3] Youda Wei, Xiaodong Hu	Title: Text Recognition from Silent Lip Movement Video Journal: IEEE Access	3D convolutional neutral networks, deep learning. First, the visual to audio feature architecture maps a variable-length sequence of video frames to the auditory MFCC features. Second, the audio feature to text architecture distinguishes the text information from the audio feature	The dataset used for training the network was the GRID audio-visual corpus which consists of audio and video recordings of 34 different speakers (male and female).	Average validation accuracy is 92.76% and for "unknown" label can be up to 91.39%.	Due to the size of the GRID dataset, the vocabulary is relatively small. The future work is to collect more train data and to propose a more robust and accurate structure to detect the lips movement in real life situation for large vocabulary dataset.
[4] Adriana Fernandez-Lopez, Federico M. sukno	Title - End-to-End Lip-Reading Without Large-Scale Data Journal - IEEE/ACM VOL. 30, 2022	Automatic Lip Reading (ALR) using Deep Neural Network (DNN)	VLRF dataset TCD-TIMIT dataset	44.77%CER and 72.90% WER 36.58% CER and 56.29% WER	Working On large scale data and also focuses on transfer learning between languages.
[5] Leyuan Qu, Cornelius Weber and Stefan Wermter	Title: LipSound2: Self-Supervised Pre-Training for Lip-to-Speech Reconstruction and Lip Reading Journal: IEEE Access	Self-supervised pre-training, speech recognition, speech reconstruction LipSound2 which consists of an encoder-decoder architecture and location aware attention mechanism to map face image sequences to Mel-scale spectrograms directly without requiring any human annotations	VoxCeleb2 is a large-scale audio-visual corpus, extracted from YouTube videos. GRID and TCD-TIMIT datasets are in controlled experimental environments with fixed frontal face angle and clean background in audio and vision. CMLR (Chinese Mandarin Lip Reading) is collected from videos by 11 hosts of the Chinese national news program.	-	Audios produced on a well pretrained speech recognition model for both English and Chinese lip-reading experiments. Future work will focus on more realistic configuration, such as the variety of light conditions, moving head poses and different background environments.

[6] PV Sindhura	Title: Convolutional Neural Networks for Predicting Words: A Lip-Reading System Journal: IEEE Access	The study uses CNNs (Inception V3 and Alex Net) to extract mouth regions from films and interpret lips to words. Tests that are speaker-independent and speaker-dependent are administered, along with the application of transfer learning.	1.Miracl-VC1 dataset (containing data from 15 speakers,10 words and phrases)	1.AlexNet: 86.6% 2.Inception V3: 64.6%	The paper highlights the need for advanced lip-reading techniques, especially with dynamic mouth contours, rich semantic information, and application of adaptive graph structures. It presents a new application of graph convolution in lip reading that addresses previous gaps in the field.
[7] Dr. Mamatha G1, Bharath Roshan B R2, Vasudha S R3	Title: Lip Reading to Text using Artificial Intelligence Journal: International Journal of Engineering Research & Technology (IJERT)	Uses Combination of a convolutional neural network (CNN) and an attention-based long short-term memory (LSTM) for lip-reading recognition.	LRW	88.2%	The proposed system may be trained and evaluated on a specific dataset, which may limit its generalizability to other languages, speech modes, and imaging conditions. Future research could explore techniques to improve the accuracy and reliability of lip-reading systems
[8] Yiting Li, Yuki Takashima, Tetsuya Takiguchi, Yasuo Ariki	Title: Lip Reading Using a Dynamic Feature of Lip Images and Convolutional Neural Networks Journal: IEEE Access	Use of Convolutional Neural Networks (CNNs) for processing the dynamic feature, reducing negative influences and also face alignment blurring.	No Mention	71.76%	The recognition of word could be done from the different angles and distance, which improves the quality of the system.
[9] Fatemeh Vakhshiteh, Farshad Almasganj	Title: lip-reading via Deep Neural Network Using Appearance-based Visual Features. Journal: IEEE Access	The Deep Belief Network (DBN) is used for the recognition part of the lip reading. The features are extracted from the images and word using Deep Neural Networks (DNN).	Cuave	45.63%	The accuracy of the system can be increased using the different techniques at the same time for preprocessing.
[10] Nergis Pervan Akman, Talya Tumer Sivrij, ali Berkol, Hamit Erdem	Title - Lip Reading Multiclass Classification by Using Dilated CNN with Turkish Dataset Journal – ICECET 2022	It is evaluated by using Dilated Convolutional Neural Network (DCNN), a different variation of CNN.	Turkish dataset	58.90	Extend the dataset and work on developing a new preprocessing strategy also the extending the dataset focusing on new phrases and words and size of

					each class will improve the results.

[11] Huijuan Wang	Title: A Lip Reading Method based on 3D Convolutional Vision Transformer Journal: IEEE Access	The methodology includes data preprocessing, front-end (3DCvT) and back-end (BiGRU) network, model variants (3DCvT-I, II, III), Adam optimizer training, and label smoothing for better cross-entropy loss. calibration.	1.LRW (English word reding dataset) 2.LRW-1000(Chinese word reading Dataset	1.LRW: 88.5% 2.LRW-1000: 57.5%	Research has identified a need for improved lip-reading methods to overcome challenges in capturing temporal and spatial information between video frames, extracting subtle features of lip movement, and addressing information loss due to resolution reduction in deep networks.
[12] Changchong Sheng	Title: Adaptive Semantic-Spatio-Temporal Graph Convolutional Network for Lip Reading Journal: IEEE	The lip-reading method improves readability with a dual stream front-end (ASST-GCN) and back-end (specialized back-end) for word and sentence work, accurately reproducing lip details, spatiotemporal face landmarks, and overcoming pre-existing limitations.	1.LRW 2.LRS2 3.LRS3	82.6%	This research addresses the challenge of dynamic lip-reading contours, the underutilization of semantic data, the introduction of adaptive graph structures, and the introduction of new graph convolution for lip reading to address an existing research gap.
[13] Alexandros Koumparoulis, Gerasimos Potamianos	Title: MobiLipNet: Resource-Efficient Deep Learning Based Lipreading Journal: ResearchGate DOI: 10.21437	MobileNet convolutional neural network Architecture for image classification. extend the 2D convolutions of MobileNets to 3D ones, in order to better model the spatio-temporal nature of the lipreading problem. visual speech recognition, deep learning, ResNet.	TCD-TIMIT corpus contains audio-visual recordings of continuous speech by 62 speakers uttering 6913 phonetically-rich TIMIT sentences (6k word vocabulary) in studio-like conditions	WER - 53.01%, This result is very close to the ResNet WER of 52.94%	MobiLipNetV2, that has 106 times less parameters and is 37 times faster than the state-of-the-art ResNet. Further, the model outperforms a baseline 3D-CNN by 4.27% absolute in WER, 12 times in computational efficiency, and 20 times in size.
[14] Kenji Matsui, Kohei Fukuyama, Yoshihisa Nakatoh, Yumiko O. Kato	Title: Speech Enhancement System Using Lip-reading Journal: IEEE Access	Word recognition experiment using VAE encoder and CNN performed with 20 Japanese words. 36 viseme images were converted into very small data using VAE (Variational Auto Encoder), then the training data for word recognition model was generated.	SSSD (Speech Scene database by Smart Device) The training data was 72 people uttered 25 words 10 times each. The test data uses 5000 sample utterance videos that are not	65%	Requires a huge amount of data and can only recognize words existing in the data set. Therefore, they investigated a method that can recognize the word that one wants to utter and optimize it for the user by using a small amount of data.

			included in the training data.		
[15] Daehyun Lee, Kyungsik Myung	Title: Read My Lips, Login to the Virtual World Journal: IEEE Access	RNN (Recurrent Neural Network) is used for processing sequential data, LSTM (Long Short-Term Memory) neural network architecture that can translate visual utterance to password as a knowledge factor. Open CV was used for extracting the face region.	GRID dataset contains 34 speakers utter 1,000 sentences with a sequence of words in the form "verb color preposition digit letter adverb".	93.8%	This paper proposes a multi-factor authentication system architecture by lip-reading and the iterative method to improve accuracy as good as state-of-the-art performance. For future work, various data should be gathered and tested under real-life conditions.
[16] Dong-Won Jang, Hong-In Kim, Changsoo Je, Rae-Hong Park, And Hyung-Min Park	Title: Lip Reading Using Committee Networks with Two Different Types of Concatenated Frame Images. Journal: IEEE Access	It uses convolutional neural network (CNN) to analyze two kinds of combined frame images (CFIs), complete lip images and smaller image patches around important lip landmarks.	OuluVS2	88.9%	The two different types of CFIs can be improved the accuracy and the performance of the system rather than using a single.
[17] Tayyip Özcan, Alper Basturk	Title: Lip Reading Using Convolutional Neural Network with and without Pre-Trained Models Journal: Balkan Journal of Electrical and Computer Engineering	Lip reading from video is performed by using the CNN technique. The standard and Av letters datasets used for training and testing the CNN.	Av Letters	64.40%	Other pre-trained models can be support to the CNN to run on Av Letters and also different datasets.
[18] Souheil Fenghour (Associate Member IEEE), Daqing Chen (Member IEEE)	Title - Lip Reading Sentences Using Deep Learning with Only Visual Cues Journal – IEEE Xplore	The classification of Visemes which is used to convert visemes to words using perplexity analysis.	LRW and LRW-1000 BBC LRS2 training set 45839 Sentences for training and 1243 Sentences for Testing.	64.06%	Efficient conversion of Visemes to words is crucial when using visemes as classification schema for lip reading sentences.
[19] Xi Ai	Title: Cross-Modal Language Modeling in Multi-Motion-Informed Context for Lip	The paper presents a new approach to lipreading using a cross-modal language model that eliminates the need for a deep encoder and emphasizes context-informed,	1.LRS2 2.LRS3 3.LRW	89.7%	A research gap relates to the need for comprehensive evaluation, including different datasets and

	Reading p Journal: IEEE Access	multi-movement for transcription from silent videos.	4.GRID		training scenarios, and further investigation of pre-training strategies in reading models.
[20] Hassan Akbari, Himani Arora, Liangliang Cao, Nima Mesgarani	Title: Lip2audspec: Speech Reconstruction from Silent Lip Movements Video	The CNN and LSTM models are used to training the system and reconstruct the speech from lip reading.	GRID	79%	Future work: collect the more trained data and also inclide the emotions in speech.
[21] Praneeth Nemani, Ghanta Sai krishna , Nikhil Ramisetty	Title - Deep Learning based Holistic Speaker Independent Visual Speech Recognition Journal – IEEE Xplore DOI 10.1109/TAI.2022 .3220190	3DCNN Architecture is used to extracting the Spatio-temporal features and mapping the elements in the croups.	MIRACL-VC1 dataset	Training accuracy – 80.02% Testing accuracy – 77.09	In future, the existing protocol can be extended to other languages like French, Hindi, and Telugu. Also, the proposed system could be improved to enhance the performance of similar-sounding words
[22] Ahsan Adeel, Mandar Gogate, Amir Hussain, and William M. Whitmer	Title: Lip-Reading Driven Deep Learning Approach for Speech Enhancement. Journal: IEEE Transactions on Emerging Topics in Computational Intelligence	A novel lip-reading driven deep learning framework is used for lip reading.	Grid and ChiME3	80%	The noise estimatipopons can be add in the system for better performance and enhance the outcome of system.

4. Algorithm Survey

Table -2: Algorithmic Survey of Research Studies

Ref.no	Paper Title	Algorithm Name	Accuracy	Time Complexity	Space Complexity
[1]	Lipread Net: A Deep Learning Approach to Lip Reading	3-Dimensional Convolutional Neural Network (3DCNN)	93%	O(O*P*Q) where O, P, Q are the input volume dimensions.	O(I*J*K*O*P*Q) where I, J, K are the filter dimensions and O, P, Q are input volume dimensions
[4]	End-to-End Lip-Reading Without Large-Scale Data	Deep Neural Network (DNN)	96 %	O(N) where N is the size of the input.	O(M) where M is the size of the input.
[10]	ip Reading Multiclass Classification by Using Dilated CNN with Turkish Dataset	lated Convolutional Neural Network (DCNN)	58.90 %	O(L*M*N) where L is the number of layers in the network, M is the number operations required per layer and N is the input samples being processed in a batch.	O(P) where P is the total number of learnable parameters in the network

[6]	Convolutional Neural Networks for Predicting Words: A Lip-Reading System	Convolutional Neural Network (CNN)	86.06%	O(N*K*M) where N is the number of input features, K is the size of the convolutional Kernel and M is the number of filters the convolutional layer.	-
[2]	Lip Reading Using Temporal Convolutional Networks.	Temporal Convolutional Networks (TCN)	85.03%	O(L*N*K) where L is the number of convolutional layers, N is the length of the input sequence and K is the size of the convolutional kernel.	-
[15]	Read My Lips, Login to the Virtual World	Recurrent Neural Network (RNN)	93.08%	O(T*D*D') where T is the number of steps in the input sequence, D is the Dimensionality of the input at each time step and D' is the number of hidden units or dimension in the RNN's hidden state.	O(D') where D' is the number of hidden units dominates the memory usage.
[7]	Lip Reading to Text using Artificial Intelligence	Combination of Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM)	88.02%	O(L*N*C*H*W*F*F') + O(T*D*D') where L is the number of layers in the network, N is the Batch size, H is the height of the input data, W is the width of the input data, F is the number of Filters in the convolutional layers, F' is the number of Filters in the Subsequent layers, T is the Sequence length or time steps D is the number of hidden states in the LSTM layer and D' is the number of LSTM units or hidden states in the subsequent LSTM layer	O(W) + O(D') where this term represents the Space complexity of fully connected layers

5. Dataset Survey

Table -3: Dataset Survey of Lip-Reading Recognition

Dataset Name	Description	Language	Year	Creator	Advantages	Disadvantages
LRW	Audio-visual speech recognition dataset collected from in-the-wild videos	English	2016	Oxford University	Large vocabulary, high-quality videos, and diverse speakers	It has limited Variability and data size.
LRS2	io-visual speech recognition dataset collected from in-the-wild videos	ish	2017	Oxford University	Large vocabulary, high-quality videos, and diverse speakers	Limited number of words
LRS3	Audio-visual speech recognition dataset collected from TED and TEDx videos	English	2019	Oxford University	Large vocabulary, high-quality videos, and diverse speakers	It has smaller size and Privacy Concerns
GRID	A corpus of 1,000 sentences spoken by 34 speakers in a controlled environment.	ish	2005	University of Sheffield	High-quality videos and audio	It consists of a single speaker, resulting limited variability.
TCD-TIMIT	set of 5,898 sentences spoken by 62 speakers in a controlled environment.		13	Trinity College Dublin	High-quality videos and audio	d number of ers
AV Letters	Audio-visual speech recognition dataset collected from a single speaker	English	2002	Jniversity of East Anglia	The quality of videos and audio is high	ted y and letters
OuluVs2	dataset of 20 short phrases spoken by 52 speakers in a controlled environment	Finnish	2011	University of Oulu	Large vocabulary and diverse speakers	The average number of samples in each class is relatively low
CUAVE	Audio-visual speech recognition dataset collected from multiple speakers	English	2004	Carnegie Mellon University	Large vocabulary and diverse speakers	Limited number of words
LRW 1000	A naturally-distributed large-scale benchmark for lip reading in the wild, which contains 1,000 classes with 718,018 samples from more than 2,000 individual speakers.	in	2016	Jniversity of Oxford	Large vocabulary, high-quality videos, and diverse speakers	Limited languages and Phoneme coverage
Lip Reading in the Wild (LRW)	A large-scale audio-visual database that contains 500 different words from over 1,000 speakers.	English	2016	versity of Oxford	Large vocabulary, high-quality videos, and diverse speakers	Limited number of

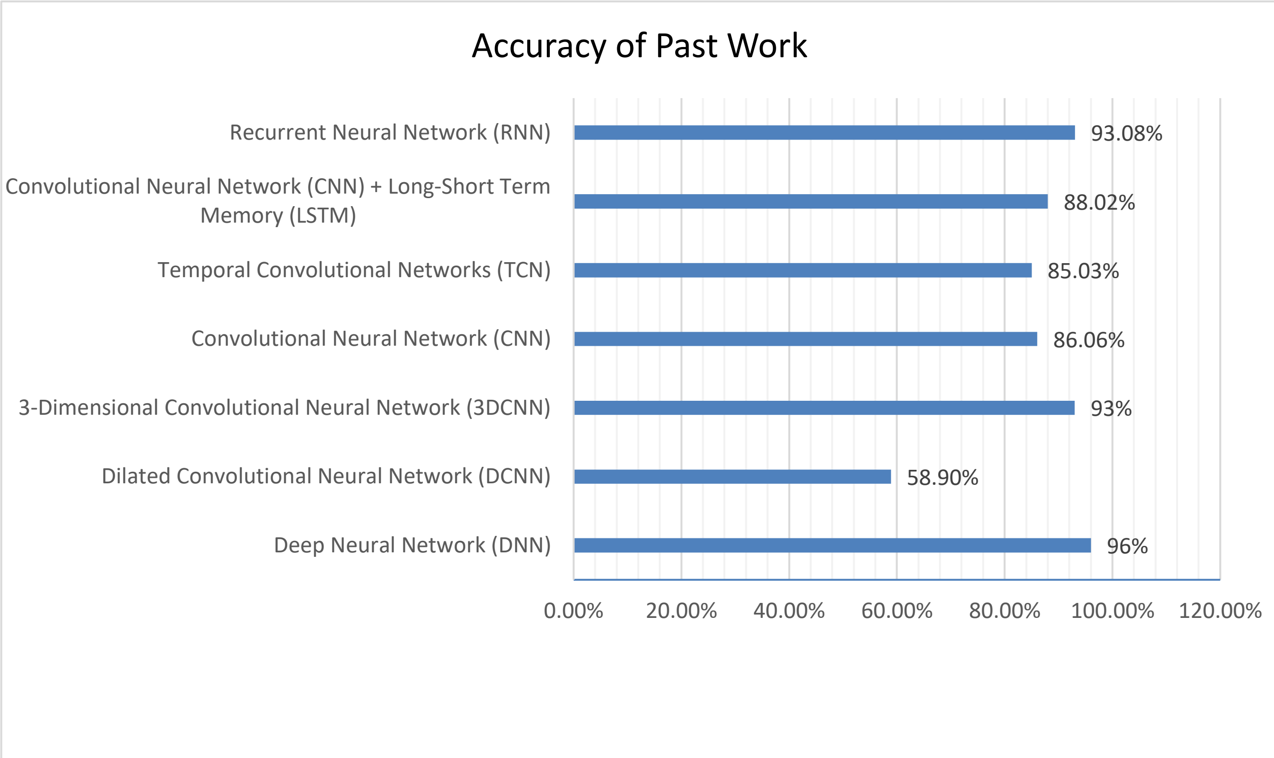

Figure 1. Graph of Accuracy Comparison on different Algorithms [1],[3],[4],[6],[10],[15]

5. CONCLUSION

In conclusion, this lip reading is currently dominated by 3D Convolutional Neural Networks (CNN), LSTM, HMM which are currently showing high accuracy [1],[6]. Numerous datasets have been examined in English and other languages, highlighting the significance of varied and extensive data for training and assessment. There are many available datasets made in various languages but we need to create more datasets in regional languages as well. The research further highlights problems in lip reading, from phoneme ambiguity to changes in lighting, head positions and the need for sufficient training data. It also explores the potential of combining visual and auditory cues to improve accuracy, particularly in noisy environments. Lip reading has the potential to revolutionize a variety of applications, from accessibility for the hearing impaired to safety and autonomous vehicles, and holds promise for a more inclusive and safer future. This survey provides us with a quick overview of the several Deep Learning techniques that can produce superior outcomes.

REFERENCES

[1] Kuldeep Vayadande, Tejas Adsare, Neeraj Agrawal, Tejas Dharmik, Aishwarya Patil, Sakshi Zod; Lipread Net: A Deep Learning Approach to Lip Reading 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC) | 979-8-3503-2379-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAISC58445.2023.102004

[2] Brais Martinez, Pingchuan Ma, Stavros Petridis, Maja Pantic; Lipreading Using Temporal Convolutional Networks 978-1-5090-6631-5/20/$31.00 ©2020 IEEE

[3] Youda Wei, Xiaodong Hu; Text Recognition from Silent Lip Movement Video 978-1-5386-6396-7/18/$31.00 ©2018 IEEE

[4] Adriana Fernandez-Lopez, Federico M. sukno; End-to-End Lip-Reading Without Large-Scale Data 2329-9290 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

[5] Leyuan Qu, Cornelius Weber and Stefan Wermter; LipSound2: Self-Supervised Pre-Training for Lip-to-Speech Reconstruction and Lip-Reading IEEE Transactions On Neural Networks And Learning Systems arXiv:2112.04748v2 [cs.SD] 12 Sep 2022.

[6] PV Sindhura; Convolutional Neural Networks for Predicting Words: A lip-Reading System 978-1-5386-5130-8/18/$31.00 ©2018 IEEE

[7] Dr. Mamatha G1, Bharath Roshan B R2, Vasudha S R3; Lip Reading to Text using Artificial Intelligence International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV9IS010312 Vol. 9 Issue 01, January-2020

[8] Yiting Li, Yuki Takashima, Tetsuya Takiguchi, Yasuo Ariki; Lip Reading Using a Dynamic Feature of Lip Images and Convolutional Neural Networks 978-1-5090-0806-3/16/$31.00 copyright 2016 IEEE

[9] Fatemeh Vakhshiteh, Farshad Almasganj; lip-reading via Deep Neural Network Using Appearance-based Visual Features 2017 24th national and 2nd International Iranian Conference on Biomedical Engineering (ICBME), Amirkabir University of Technology, Tehran, Iran, 30 November - 1 December 2017 978-1-5386-3609-1/17/$31.00 ©2017 IEEE

[10] Nergis Pervan Akman, Talya Tumer Sivrij, ali Berkol, Hamit Erdem; Lip Reading Multiclass Classification by UsingDilated CNN with Turkish Dataset 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET) | 978-1-6654-7087-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICECET55527.2022.9873011

[11] Huijuan Wang. GANGQIANG PU, AND TINGYU CHEN; A Lip Reading Method Based on 3D Convolutional Vision Transformer Digital Object Identifier 10.1109/ACCESS.2022.3193231 VOLUME 10, 2022

[12] Changchong Sheng Xinzhong Zhu , Member, IEEE, Huiying Xu , Member, IEEE,Matti Pietikäinen , Fellow, IEEE, and Li Liu , Senior Member, IEEE; Adaptive Semantic-Spatio-Temporal Graph Convolutional Network for Lip Reading 1520-9210 © 2021 IEEE IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 24, 2022

[13] Alexandros Koumparoulis, Gerasimos Potamianos; MobiLipNet: Resource-Efficient Deep Learning Based Lipreading DOI: 10.21437/Interspeech.2019-2618 September 15–19, 2019, Graz, Austria

[14] Kenji Matsui, Kohei Fukuyama, Yoshihisa Nakatoh, Yumiko O. Kato; Speech Enhancement System Using Lip-reading 978-1-7281-6946-0/20/$31.00 ©2020 IEEE.

[15] Daehyun Lee, Kyungsik Myung; Read My Lips, Login to the Virtual World 978-1-5090-5544-9/17/$31.00 ©2017 IEEE 2017 IEEE International Conference on Consumer Electronics (ICCE)

[16] Dong-Won Jang, Hong-In Kim, Changsoo Je, Rae-Hong Park, And Hyung-Min Park; Lip Reading Using Committee Networks with Two Different Types of Concatenated Frame Images Digital Object Identifier 10.1109/ACCESS.2019.2927166 VOLUME 7, 2019

[17] Tayyip Özcan, Alper Basturk; Lip Reading Using Convolutional Neural Networks with and without Pre-Trained Models Article in Balkan Journal of Electrical and Computer Engineering · April 2019 DOI: 10.17694/bajece.479891

[18] Souheil Fenghour (Associate Member IEEE), Daqing Chen (Member IEEE); Lip Reading Sentences Using Deep Learning with Only Visual Cues Digital Object Identifier 10.1109/ACCESS.2020.3040906 VOLUME 8, 2020.

[19] Xi Ai and Bin Fang; Cross-Modal Language Modeling in Multi-Motion-Informed Context for Lip Reading IEEE/ACM Transactions on Audio, Speech, And Language Processing, VOL. 31, 2023 2329-9290 © 2023 IEEE.

[20] Hassan Akbari, Himani Arora, Liangliang Cao, Nima Mesgarani; Lip2audspec: Speech Reconstruction from Silent Lip Movements Video 978-1-5386-4658-8/18/$31.00 ©2018 IEEE.

[21] Praneeth Nemani, Ghanta Sai krishna , Nikhil Ramisetty; Deep Learning based Holistic Speaker Independent Visual Speech Recognition Citation information: DOI 10.1109/TAI.2022.3220190 © 2022 IEEE.

[22] Ahsan Adeel, Mandar Gogate, Amir Hussain, and William M. Whitmer; Lip-Reading Driven Deep Learning Approach for Speech Enhancement 2471-285X © 2019 IEEE IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, VOL. 5, NO. 3, JUNE 2021

BIOGRAPHIES

Mr. Shubham Kalburgi
Project Research Fellow and B.E. Student at Department of Computer Engineering, RMD Sinhgad School of Engineering,SPPU, Pune.

Prof. Vina M. Lomte
Project Guide and Head of Computer Engineering Department at RMD Sinhgad School of Engineering, SPPU, Pune

Mr. Niraj Jain
Project Research Fellow and B.E. Student at Department of Computer Engineering, RMD Sinhgad School of Engineering,SPPU, Pune.

Mr. Ashutosh Buge
Project Research Fellow and B.E. Student at Department of Computer Engineering, RMD Sinhgad School of Engineering,SPPU, Pune.

Ms. Shreya Kankal
Project Research Fellow and B.E. Student at Department of Computer Engineering, RMD Sinhgad School of Engineering,
SPPU, Pune.