Using Deep Learning in Sign Language Translation to Text

Mary Jane C. Samonte
Doctor in Information Technology
Faculty Member of School of Information Technology
Mapua University, Makati City, Philippines
mjcsamonte@yahoo.com

Carl Jose M. Guingab, Ron Andrew Relayo,
Mark Joseph C. Sheng, John Ray D. Tamayo
School of Information Technology, Mapua University
Makati City, Philippines, cjmguingab@mymail.mapua.edu.ph,
rarelayo@mymail.mapua.edu.ph, mjcsheng@mymail.mapua.edu.ph,
jrdtamayo@mymail.mapua.edu.ph

Abstract

This study is about sign language translation into text; this technology is for those experiencing difficulty in communicating, i.e., those who have speech disorders, hearing impairment, and are deaf. Sign language translation into text is done through a process when one uses sign language that could be translated into text. Deep learning is a machine learning and artificial intelligence (AI) technique that mimics how humans acquire knowledge; the complexity and abstraction of deep learning algorithms are stacked in a hierarchy. The researchers utilized Convolutional Neural Networks (CNN), Connectionist Temporal Classification (CTC), and Deep Belief Network (DBN) to collect and gather the results of their system through different methods of translating sign language into text. Through the systematic literature review of various published papers related to sign language translation using deep learning techniques, this paper discusses and systematically analyzed various studies and come up with a proposed model in translating sign language into text.

Keywords
Sign Language Recognition, Systematic Literature Review, Deep Learning, Natural Language Processing

1. Introduction
The study is about sign language translation into text; this technology is for those who are experiencing difficulty in communicating i.e., those who have speech disorders, hearing impairment, and are deaf (Fang et al, 2017) (Zhang et al, 2019) (Ferreira et al, 2019) (Sharma et al, 2021) (Sridhar et al, 2020) (Alam et al, 2021) (Orbay and Akarun, 2020) (Shurid et al, 2020) (Hou et al, 2019)]. Sign language translation into text will help them communicate with other people as they can express their feelings thoroughly without the concern of misunderstanding them, or not understanding them at all. Moreover, the problem with the difficulty in communication and comprehension is prevalent all over the world, which elaborates the need for sign language translation into text.

This research seeks to focus on related articles and studies that tackled deep learning in sign language translation into text, which fluctuated through the years 2017 to 2021. There were articles that did not make use of deep learning or NLP, which will not be included in the scope of this paper. Specifically, important details such as algorithms used in the study will be discussed. Alongside this, the different sources of input in the study and the type of content that the research papers have used to recognize sign language will be discussed. With the information gathered, a proposed system will be discussed.

2. Methods

The following are the internet databases that were utilized in this study: IEEE Xplore, Elsevier, SpringerLink, Taylor and Francis online, and World Scientific. A number of procedures are taken when seeking publications in the aforementioned online databases: (1) Type a term into an online database; (2) Filter the publication date to only see results from 2017 to 2021; (3) Sort the studies into the discovered studies document; (4) Sort the studies into the filtered studies document based on the title, abstract, and results. (5) Put only studies in the chosen studies column that are relevant to the investigation.

3.1 Study Filtering

There are a total of forty (40) papers obtained through the selection process; however, after filtering, only twenty (20) articles remain which fulfill the requirements based on their abstract and title. With a total of twenty (20) studies, IEEE Xplorer = 6, ACM = 6, Springer = 4, Elsevier = 2, Taylor and Francis Online = 1, and World Scientific = 1 were chosen and screened. The studies were chosen for their importance to deep learning, sign language recognition, and translation.

The chosen studies were determined after screening the articles. The studies that satisfy the criteria for searching based on their relevance to the issue account for the small number of research available. The summary of the details of each study in sign language translation is shown in Table 1. Specifically, the table enumerates the different publishers of the paper, which are reputable sources such as ACM, Springer, Elsevier, and more. The different deep learning techniques are also listed, some of which uses a variation of neural networks, LTSM, and other classifier techniques. Since there are multiple sign languages used as input to these studies, they are also included in the table below.

Table 1. Sign Language Translated

Title of the Study	Source of Study	Features and Classifiers	Sign Language Used
Sign Language Fingerspelling Recognition Using Depth Information and Deep Belief Networks	World Scientific	Deep-Belief Network (DBN)	American Sign Language (ASL)
DeepASL: Enabling Ubiquitous and Non-Intrusive Word and Sentence-Level Sign Language Translation	ACM	Hierarchical bidirectional deep recurrent neural network (HB-RNN) + Connectionist Temporal Classification (CTC)	American Sign Language (ASL)
Deep learning-based Sign language recognition system for static signs	Springer	Convolutional Neural Networks (CNN)	Indian Sign Language (ISL)
Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM	Elsevier	Convolutional Neural Networks (CNN) + Bidirectional LSTM	Arabic Sign Language (ArSL)
MyoSign: enabling end-to-end sign language recognition with wearables	ACM	Convolutional Neural Networks (CNN) + Bidirectional Long Short Term Memory Layers (BLSTM)	American Sign Language (ASL)

Enabling Real-time Sign Language Translation on Mobile Platforms with On-board Depth Cameras	ACM	3-dimensional convolutional neural network (3DCNN)	Korean Sign Language (KSL)
SonicASL: An Acoustic-based Sign Language Gesture Recognizer Using Earphones	ACM	SubNet + Connectionist Temporal Classification (CTC)	American Sign Language (ASL)
DeSIRe: Deep Signer-Invariant Representations for Sign Language Recognition	IEEE	Convolutional Neural Networks (CNN)	Portuguese Sign Language (PSL) + American Sign Language (ASL)
Exploration of cChinese sign language recognition using wearable sensors based on deep belief net	IEEE	Deep-Belief Net (DBN)	Chinese Sign Language (CSL)
Benchmarking deep neural network approaches for Indian Sign Language recognition	Springer	Deep Convolutional Neural Network (DCNN)	Indian Sign Language (ISL)
Connectionist Temporal Fusion for Sign Language Translation	ACM	Connectionist Temporal Classification (CTC)	German Sign Language (DGS) + Chinese Sign Language (CSL)
SubUNets: End-to-End Hand Shape and Continuous Sign Language Recognition	IEEE	Convolutional Neural Networks (CNN) + Bidirectional Long Short Term Memory Layers (BLSTM) + Connectionist Temporal Classification (CTC)	German Sign Language (DGS)
INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition	ACM	Bidirectional Long Short Term Memory Layers (BLSTM) + 3-dimensional convolutional neural network (3DCNN)	Indian Sign Language (ISL) + American Sign Language (ASL) + Danish Sign Language (DSL)
Two Dimensional Convolutional Neural Network Approach for Real-Time Bangla Sign Language Characters Recognition and Translation	Springer	Convolutional Neural Network (CNN)	Bangla Sign Language (BdSL)
Vision-based hand gesture recognition for indian sign language using convolution neural network	Taylor and Francis Online	Convolutional Neural Network (CNN)	Indian Sign Language (ISL)
Deep Learning-Based Sign Language Digits Recognition	IEEE	Convolutional Neural Network (CNN)	American Sign Language (ASL)

From Thermal Images With Edge Computing System			
Vision-based hand gesture recognition using deep learning for the interpretation of sign language	Elsevier	Convolutional Neural Network (CNN)	Indian Sign Language (ISL) + American Sign Language (ASL)
Neural sign language translation by learning tokenization	IEEE	3-dimensional convolutional neural network (3DCNN) + 2-dimensional convolutional neural network (2DCNN)	German Sign Language (DGS)
Bangla sign language recognition and sentence building using Deep Learning	IEEE	Convolutional Neural Network (CNN)	Bangla Sign Language (BdSL)
SignSpeaker: A Real-time, High-Precision SmartWatch-based Sign Language Translator	ACM	Connectionist Temporal Classification (CTC)	American Sign Language (ASL)

4. Literature Review

This section cites a variety of journal (Hu et al, 2018) (Yu et al, 2019) (Sharma et al, 2021) (Gangrade and Bharti, 2020) (Breland et al, 2021) (Sharma and Singh, 2021) and conference (Fang et al, 2017) (Zhang et al, 2019) (Cihan Camgoz et al, 2017) (Sridhar et al, 2020) (Orbay and Akarun, 2020) (Shurid et al, 2020) (Hou et al, 2019) articles to show how they appear in the references section. It also cites different research papers (Wadhawan and Kumar, 2020) (Alam et al, 2021) and proposal studies (Abdul et al, 2021) (Park et al, 2021) (Jin et al, 2021) (Ferreira et al, 2019) (Wang et al, 2018). The goal of this paper is to identify the different methods used to translate sign language to text to understand people with disabilities, especially the deaf community. The different creators and innovators used (1) sign language recognition and (2) sign language to text or sign language translation method in their study. The different gathered and filtered studies clustered by publication year with 2019 and 2021 as the highest number of listed used studies in this paper. Year 2017 (Three (3) with reference number of (Hu et al, 2018) (Fang et al, 2017)); year 2018 (One (1) with reference number of (Wang et al, 2018)) and year 2020 (Four (4) with reference number of (Sharma et al, 2021) (Sridhar et al, 2020) (Gangrade and Bharti, 2020) (Orbay and Akarun, 2020)).

4.1 Sign Language Recognition

The sign language recognition accuracy depends on the focused system, and mostly they have a high-level accuracy rate. Sign language recognition is one most used system because of its accuracy and validation rate. It shows many studies about sign language recognition because it is most commonly used in translating sign language, which gives a high accuracy rate depending on the dataset available. The process of deep learning models in translating sign language to text in the translation based on the gathered and selected studies is shown in Figure 1. The various studies under sign language recognition containing its content, description and respective references are shown in Table 2.

Table 2. Different studies with sign language recognition

Content	Description	Reference
Hand Gesture	Through the use of hand gesture to sign language recognition	(Jin et al, 2021), (Ferreira et al, 2019), (Sharma et al, 2021), (Cihan Camgoz et al, 2017), (Gangrade and Bharti, 2020), (Sharma and Singh, 2021) (Shurid et al, 2020)

Web Camera Based	Web camera based on dataset of images and camera itself	(Abdul et al, 2021), (Sridhar et al, 2020), (Alam et al, 2021), (Breland et al, 2021), (Shurid et al, 2020)
Videos	Videos as data sets to recognized sign language	(Abdul et al, 2021), (Wadhawan and Kumar, 2020), (Yu et al, 2019), (Cihan Camgoz et al, 2017), (Sridhar et al, 2020), (Alam et al, 2021)
Wearable sensors	The wearable end-to-end gadget for sign language recognition	(Jin et al, 2021), (Yu et al, 2019)

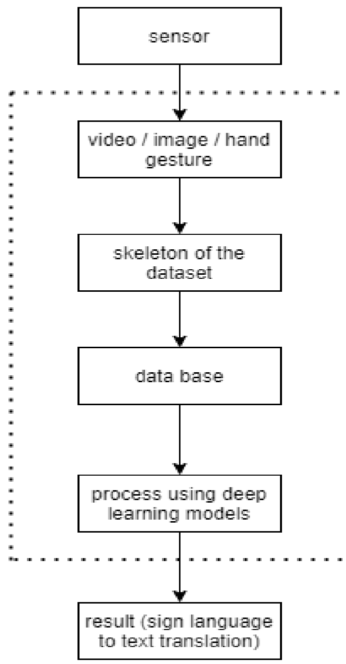

Figure 1. This shows the process of deep learning models in translating sign language to text in the translation based on the gathered and selected studies.

4.2 Sign Language-to-text / Sign Language Translation

Sign language translation is another system made by the researchers to understand sign language better and get the highest accuracy translation rate. It compromises the different methods using various datasets available (images). The result of sign language translation depends on the execution of the sign language and the quality of the captured video or images to be translated to text. The different studies with sign language translation are shown in Table 3.

Table 3. Different studies with sign language translation

Content	Description	Reference
Web Camera Based	Web camera based on dataset of images and camera itself for sign language translation	(Park et al, 2021), (Alam et al, 2021)

Videos	Videos as data sets to recognized and translate sign language	(Park et al, 2021), (Wang et al, 2018), (Alam et al, 2021), (Orbay and Akarun, 2020)
Wearable sensors	The wearable end-to-end gadgets and sensors t for sign language translation	(Fang et al, 2017), (Hou et al, 2019)

4.3 Methods and Tools Used

The study used different methods and tools such as Convolutional Neural Networks (CNN), Connectionist Temporal Classification (CTC), and Deep Belief Network (DBN) to gather the result of their system through the different methods to translate sign language to text. Out of twenty (20) papers gathered and selected, CNN has the highest percentage usage with 70% usage method (it includes DCNN (Sharma et al, 2021), 2DCNN (Orbay and Akarun, 2020), 3DCNN (Park et al, 2021), (Sridhar et al, 2020), (Orbay and Akarun, 2020) and combinations of other methods with CNN such as Bidirectional Long Short Term Memory Layers (BLSTM) (Abdul et al, 2021), (Zhang et al, 2019), (Cihan Camgoz et al, 2017), and Connectionist Temporal Classification (CTC)), followed by CTC with 20% usage (it includes combinations of other methods with CTC such as HB-RNN (Fang et al, 2017)) and DBN with 10% usage in the gathered and filtered papers.

Convolutional Neural Networks (CNN) - The most used method and one of the deep learning models in the studies. It is used by (Wadhawan and Kumar, 2020), (Abdul et al, 2021), (Zhang et al, 2019), (Park et al, 2021), (Ferreira et al, 2019), (Sharma et al, 2021), (Cihan Camgoz et al, 2017), (Sridhar et al, 2020), (Alam et al, 2021(Gangrade and Bharti, 2020), (Breland et al, 2021), (Sharma and Singh, 2021), (Orbay and Akarun, 2020) and (Shurid et al, 2020). CNN is is a tool for extracting information from images (Park et al, 2021), used in image processing and recognition (Alam et al, 2021) and composed of convolutional layers followed by additional layers (Wadhawan and Kumar, 2020).

Connectionist Temporal Classification (CTC) - The second most used method in the studies gathered in this paper. It is used by (Fang et al, 2017), (Jin et al, 2021(Wang et al, 2018), (Cihan Camgoz et al, 2017), and (Hou et al, 2019). CTC is proposed by Graves et al., one of the most often used methods for developing a sequence-to-sequence model [12] and compares between speech and optical character recognition tasks (Jin et al, 2021).

Deep Belief Network (DBN) - the least used method in the studies gathered in this paper. It is used by (Hu et al, 2018) and (Yu et al, 2019). DBN was proposed by Hinton in the year 2006, and became a breakthrough in deep learning because of its ability to overcome the flaws of traditional algorithms and a deep neural network (DNN) model made up of Restricted Boltzmann Machine (RBM) stacked together, which has captivated the interest of various researchers (Hu et al, 2018) (Yu et al, 2019).

Hierarchical Bidirectional Deep Recurrent Neural Network (HB-RNN) - Another method used along with CNN. It is used by B. Fang, J. Co, and M. Zhang and design for single-sign language to word-level ASL translation (Fang et al, 2017).

Bidirectional Long Short Term Memory Layers (BLSTM) - Another method used along with CNN and CTC. It is used by (Abdul et al, 2021) and (Zhang et al, 2019). It infers at each clip using both prior and previous information, allowing it to deal with the problem of numerous indicators having extremely identical properties at first (Zhang et al, 2019).

3-Dimensional Convolutional Neural Network (3DCNN) and 2-Dimensional Convolutional Neural Network (2DCNN) - Another method used along with CNN and BLSTM. It is used by (Park et al, 2021), (Sridhar et al, 2020) and (Orbay and Akarun, 2020). 3DCNN is used for extraction and word categorization, it uses its baseline inference model that allows for accurate sign language translation by extracting temporal patterns of a user's hand gestures from depth picture frames (Park et al, 2021).

4.4 Results Observed

Based on the different results gathered from the selected papers using deep learning methods such as CNN, BLSTM, DBN, and CTS, it obtains a high-level accuracy and rate regardless of different methods applied in the datasets (i.e., videos). Researchers assumed that the accuracy level or rate depends on the datasets used in the different deep learning methods. A high accuracy rate is possible when it has a precise execution of sign language applying those mentioned methods since most of the system depends on the dataset available, whether it is a real-time execution or not.

5. Proposed Sign Language Translation into Text System

The proposed sign language translation into text solution employs the definition of each sign via sign language recognition (Abdul et al, 2021) to yield higher accuracy in sign language translation. To translate a sign language into a text it must be recognized first. CTC has been recognized as a critical technique to separate words (Hou et al, 2019) to be able to build a sentence-level structure of translation. To be able to recognize and accurately translate a sentence-level structure of text, it is necessary to use Bidirectional Long Short-Term Memory (BiLSTM) (Wang et al, 2018). As the proposed model focuses only on the translation of sign language into text the use of BILSTM is not a requisite. None of the sign language translation literature used Natural Language Processing (NLP) features and techniques as their primary tool. However, NLP has been recognized by the study made by H. Park, Y. Lee, and J. Ko (Park et al, 2021) as an essential tool in confirming that their model can be used for accurate word-level categorization, detecting the border between successive word-level signals. Additionally, there is a distinct lack of sign language translation into text in the field of deep learning. The proposed model as a product of systematic literature review in this paper is shown in Figure 2.

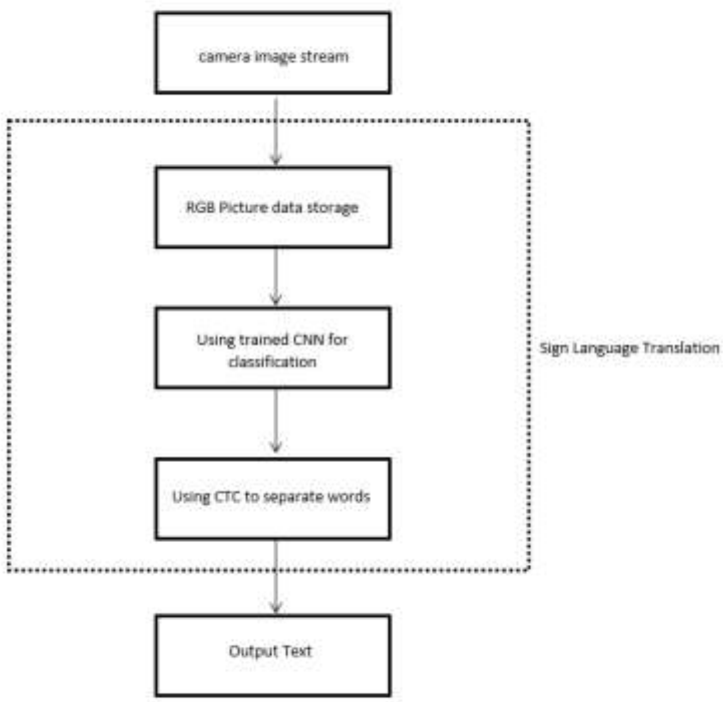

Figure 2. Proposed Sign Language Translation System using Connectionist Temporal Classification (CTC) and Convolutional Neural Network (CNN).

5. Implication and Conclusion

The systematic review paper successfully determined the factors that are present in most sign language to text methods, focusing on deep learning features and its success in identifying human gesture input and outputting accurate translations. These results were used as a basis in the development of a sign language translation to text system. The research garnered results by narrowing down the 40 selected related literature and identifying 20 relevant studies that are more relevant to deep learning using a two-step filtering process. Out of the methods being utilized, the highest one being used in most studies is CNN, with 70% usage, followed by CTC, with 20% usage, and lastly DBN, with 10% usage. The research findings of this review can be used as a reference in research papers that tackle sign language recognition and translation using deep learning methods. Additionally, the results could also be a basis for making a sign language recognition model and further encourage the development of improved sign language recognition systems.

6. Limitation and Future Research

The limitations of the research will be for any similar studies about deep learning and sign language recognition that will be conducted in the future. The studies gathered in this paper are adequate to create a sign language to text system, but for any future research that will want to delve deeper into improving the methods for sign language translation, more studies and related literature should be examined. While a proposed system was given in this paper, it is far from being usable in actual scenarios, as it is more of a concept with proposed methods used in a vacuum. Additionally, the study recommends studying further into the less utilized methods in sign language recognition, like DNN and CTC.

References

Abdul, W., Alsulaiman, M., Amin, S.U., Faisal, M., Muhammad, G., Albogamy, F.R., Bencherif, M.A. and Ghaleb, H., Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM. Computers & Electrical Engineering, vol. 95, p.107395, 2021.

Alam, M., Tanvir, M., Saha, D.K. and Das, S.K., Two Dimensional Convolutional Neural Network Approach for Real-Time Bangla Sign Language Characters Recognition and Translation. SN Computer Science, vol. 2(5), pp.1-13, 2021.

Breland, D.S., Skriubakken, S.B., Dayal, A., Jha, A., Yalavarthy, P.K. and Cenkeramaddi, L.R., Deep Learning-Based Sign Language Digits Recognition From Thermal Images With Edge Computing System. IEEE Sensors Journal, vol. 21(9), pp.10445-10453, 2021.

Cihan Camgoz, N., Hadfield, S., Koller, O. and Bowden, R., Subunets: End-to-end hand shape and continuous sign language recognition. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3056-3065), 2017.

Fang, B., Co, J. and Zhang, M., November. Deepasl: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation. In Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems (pp. 1-13), 2017.

Ferreira, P.M., Pernes, D., Rebelo, A. and Cardoso, J.S., Desire: Deep signer-invariant representations for sign language recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019.

Gangrade, J. and Bharti, J., Vision-based Hand Gesture Recognition for Indian Sign Language Using Convolution Neural Network. IETE Journal of Research, pp.1-10, 2020.

Hou, J., Li, X.Y., Zhu, P., Wang, Z., Wang, Y., Qian, J. and Yang, P., 2019, August. Signspeaker: A real-time, high-precision smartwatch-based sign language translator. In The 25th Annual International Conference on Mobile Computing and Networking (pp. 1-15), 2019.

Hu, Y., Zhao, H.F. and Wang, Z.G., Sign language fingerspelling recognition using depth information and deep belief networks. International Journal of Pattern Recognition and Artificial Intelligence, vol. 32(06), p.1850018, 2018.

Jin, Y., Gao, Y., Zhu, Y., Wang, W., Li, J., Choi, S., Li, Z., Chauhan, J., Dey, A.K. and Jin, Z., Sonicasl: An acoustic-based sign language gesture recognizer using earphones. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5(2), pp.1-30, 2021.

Orbay, A. and Akarun, L., November. Neural sign language translation by learning tokenization. In 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (pp. 222-228). IEEE, 2020.

Park, H., Lee, Y. and Ko, J., 2021. Enabling Real-time Sign Language Translation on Mobile Platforms with On-board Depth Cameras. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5(2), pp.1-30, 2021.

Sharma, A., Sharma, N., Saxena, Y., Singh, A. and Sadhya, D., Benchmarking deep neural network approaches for Indian Sign Language recognition. Neural Computing and Applications, vol. 33(12), pp.6685-6696, 2021.

Sharma, S. and Singh, S., Vision-based hand gesture recognition using deep learning for the interpretation of sign language. Expert Systems with Applications, vol. 182, p.115657, 2021.

Shurid, S.A., Amin, K.H., Mirbahar, M.S., Karmaker, D., Mahtab, M.T., Khan, F.T., Alam, M.G.R. and Alam, M.A., December. Bangla Sign Language Recognition and Sentence Building Using Deep Learning. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1-9). IEEE, 2020.

Sridhar, A., Ganesan, R.G., Kumar, P. and Khapra, M., Include: A large scale dataset for indian sign language recognition. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 1366-1375), October 2020.

Wadhawan, A. and Kumar, P., Deep learning-based sign language recognition system for static signs. Neural Computing and Applications, 32(12), pp.7957-7968, 2020.

Wang, S., Guo, D., Zhou, W.G., Zha, Z.J. and Wang, M., Connectionist temporal fusion for sign language translation. In Proceedings of the 26th ACM international conference on Multimedia (pp. 1483-1491), October 2018.

Yu, Y., Chen, X., Cao, S., Zhang, X. and Chen, X., Exploration of Chinese sign language recognition using wearable sensors based on deep belief net. IEEE journal of biomedical and health informatics, 24(5), pp.1310-1320, 2019.

Zhang, Q., Wang, D., Zhao, R. and Yu, Y., March. MyoSign: enabling end-to-end sign language recognition with wearables. In Proceedings of the 24th international conference on intelligent user interfaces (pp. 650-660), 2019.

Biographies

Mary Jane C. Samonte has a double bachelor's degree in computer education and information technology. She also has two post graduate degree; Information Technology and Computer Science. She finished her Doctor in IT with a study focusing in Deep Learning. She has a wide range of research interests that are centered around educational technologies, gamification, mobile and ubiquitous learning, digital game-based learning, artificial intelligence in education, e-health, assistive technology, natural language processing, green computing and data analytics-based studies.

Carl Jose M. Guingab is a second-year Bachelor of Science in Information Technology student of Mapua University.

Ron Andrew Relayo is a second-year Bachelor of Science in Information Technology student of Mapua University.

Mark Joseph C. Sheng is a second-year Bachelor of Science in Information Technology student of Mapua University.

John Ray D. Tamayo is a second-year Bachelor of Science in Information Technology student of Mapua University.