Received: 22 November 2022, Accepted: 23 December 2022, Published: 31 December 2022, Publisher: UTP Press, Creative Commons: CC BY 4.0

REAL TIME MALAYSIAN SIGN LANGUAGE TRANSLATOR – TUTURJOM

Arsh Obeidy

Department of Science and Information Technology, Universiti Teknologi PETRONAS, Malaysia

Email: arsh.obeidy@gmail.com

ABSTRACT

Malaysian Sign Language (MSL) is a vital language that is used as the primary means of communication between the deaf and mute and hearing communities in Malaysia. Malaysians are currently very unfamiliar with MSL, and the present platform for learning the language is scarce, not to mention the lack of efficacy of the available mobile and web learning tools. Therefore, the aim of this study is to propose software for real-time translation of MSL through a Long Short-Term Memory (LSTM) model. The paper explains six phases of CRISP-DM in creating the application while looking through various findings and important algorithms. This paper also intends to provide a clear understanding of hand gesture and pose recognition, LSTM networks and its uses in understanding sign languages.

Keywords: *LSTM, Malaysian sign language, CRISP-DM, gesture recognition, sign language recognition*

INTRODUCTION

Sign language is a way of communication that involves the use of the hands and other parts of the body [1]. According to Harini et al. [2] sign languages are natural, with their own grammar and lexicon. As such, many countries have developed their own sign languages, influenced by their environment and people's experience. MSL is used as the main mode of communication between deaf people in Malaysia [3]. With the foundation of the Malaysian Federation of the Deaf in 1998, MSL originated, and its use has grown among deaf organisations [4].

In Malaysia, deaf children make up a small percentage of the student population. Only around 20% of Deaf children are listed with the Social Welfare Department, indicating that only about 20% of Deaf children attend school [5]. Even though Malaysian Sign Language is commonly used among the deaf-mute community, people who can hear have little or no exposure to it due to a lack of a ubiquitous environment for its use [6]. The above literature shows that the accessibility of MSL is very scarce to the general public, Malaysians are unfamiliar with the MSL, and the present system for learning sign language is ineffective. As a result, there is a communication gap between the deaf and mute groups and the hearing community.

LITERATURE REVIEW

To recognize Malaysian Sign Language in real-time, this study would use Gesture recognition which is a computer technique that uses mathematical algorithms to identify and understand human gestures [7]. Gesture recognition isn't just for human hand motions; it can also be used to identify everything from head nods to various walking gaits [8]. Many applications, such as sign language recognition, augmented reality (virtual reality), and sign language interpreters for the deaf, rely heavily on hand gesture recognition [9].

There are two primary techniques for sign language recognition: image-based and sensor-based. However, a lot of research is being done on image-based techniques due to the benefit of not having to wear complicated gear like hand gloves [10], helmets, and so on, as with sensor-based approaches.

Image-based sign recognition is divided into two phases: detection and identification of signs. Sign detection is the process of extracting a feature from an item based on a specific characteristic while recognizing a certain shape that distinguishes an object from the other forms is known as sign recognition [11]. Moreover, two main methods are used for detecting hands and pose in the image, both of which are equally popular. The first method is hand detection through skin colour. The skin colour can be used to discriminate the hand region from the other moving objects in the background. The colour of the skin is measured with the HSV model. The HSV (hue, saturation, and value) value of the skin colour is 315, 94, and 37, respectively. The image of the detected hand is then resized to make the gesture recognition invariant to the image scale [12].

The second method is Human action recognition which is a subset of video content assessment that seeks to recognize actions from a set of observations on the behaviours of humans and their surroundings [13]. Finally, several Algorithms have been created based on computer vision approaches to recognize hands using various types of cameras. Skin colour, appearance, motion, skeleton, depth, and other features are all attempted to be segmented and detected by the algorithms. One of these learning methods is LSTM.

LSTM networks are recursive neural networks that can learn order dependency in sequence classification challenges. This is a necessary behaviour in complicated problem areas such as translation software, voice recognition, and also gesture recognition. LSTMs' success stems from its ability to be one of the first tools to conquer technical challenges and deliver on the promise of recurrent neural networks [14]. LSTM networks have been used for gesture recognition in many studies. For example, Hakim et al. [15] solved the challenge of dynamic gesture detection by a combination of 3DCNN and LSTM models that could extract the spatiotemporal aspects of the gesture sequence. When used in the real-time application, the FSM controller model was used to reduce the size of the gesture categorization search on the model to produce more accurate results. The results showed that, for eight selected motions, the combination of depth and RGB data had a 97.8% accuracy rate, while the FSM increased the identification rate in real-time from 89% to 91%.

In Another study done by Ameur et al. [16], LSTM model versions are used in a dynamic hand gesture recognition system that uses sequential data taken with the LMC. The research started with independently testing the fundamental U-LSTM and Bi-LSTM. Second, the D-LSTM network with various layer counts was observed. In order to create the final prediction network, known as the HBU-LSTM, a deep LSTM architecture as a hybrid model was proposed which combined the Bi-LSTM and the U-LSTM with extra components. Overall, the accuracy rates for all classes were in the range of 72.5% to 96.5%.

Lastly, Dynamic Hand Gesture Recognition (DHGR) is done by a Convolution Neural Network (CNN) and LSTM based model [17]. This system was taught to process the series of 3D input data while also learning the locations and velocities from the Leap Motion Controller (LMC). The model gave a high accuracy score of 98% compared to other DHGR techniques.

METHODOLOGY

This section will discuss the tools and the methodology used for doing this study. The rationale behind the particular process will be explained with various concepts and algorithms used to build the machine learning model for MSL recognition.

The study would be using CRISP-DM as its main framework. The CRISP-DM (Cross Industry Standard Process for Data Mining) is a six-phase model developed to represent the machine learning life cycle accurately. It is one of the most common practices in the industry for building machine learning models. The six phases of the model are explained in detail below.

Business Understanding

The first step in the CRISP-DM process is to understand what is required to achieve in terms of business. The objective of this stage of the process is to identify key elements that may have an impact on the study's result [18]. This stage also helps in finding the primary objectives of the study, entailing explaining the key goal from a business standpoint. It should be noted that after the above research, the main objective that was decided for this study would be developing an application which translates MSL in real-time through a camera.

Data Understanding and Preparation

The data specified in the study resources must be acquired in the second stage of the CRISP-DM procedure. Understanding the data is essential in creating an efficient machine-learning model [19]. This section describes the data resources gathered from different locations and the data format used.

Since the study requires the interpretation of MSL through a camera. Hence, in order to train a machine learning model that could recognize MSL, there is a need for a dataset of images or videos of people signing in MSL, coupled with what English or Malay word each photo represents. This kind of data exists for American Sign Language (ASL), but as it turns out, there is no proper dataset for MSL. For this problem, the two solutions were to either continue to research more on existing MSL databases or combine an ASL database with a database originally created by the author (The videos for this were collected through a webcam using Open CV). The second method was able to succeed because MSL is highly based on ASL. Consequently, the end database was a mixture of different basic MSL signs, with each sign having its own folder, which had 30 videos for that particular sign, and each video had 30 frames. These videos include alphabet and number signs and basic words needed for daily conversations.

Next, data preparation will take place, which entails bringing the data quality up to the standards required by the analytic methodologies [20]. The media pipe holistic was used to determine various landmarks and key points on the face and hands, which will ultimately aid in building the machine learning model. In the first step, images are collected using Open CV, which essentially uses BGR as the colour channel, but to determine key points through media pipe holistic, an image in RGB channel was needed. Therefore, as shown in Figure 1, the images were converted from BGR to RGB and then put through a media pipe model to extract the key points.

As mentioned above, media pipe holistic can be used to determine various key points from the face, pose and hands. An example is shown in Figure 2 for a frame where key points are detected and displayed. The landmark's styling was also formatted for the thickness and colors of the lines.

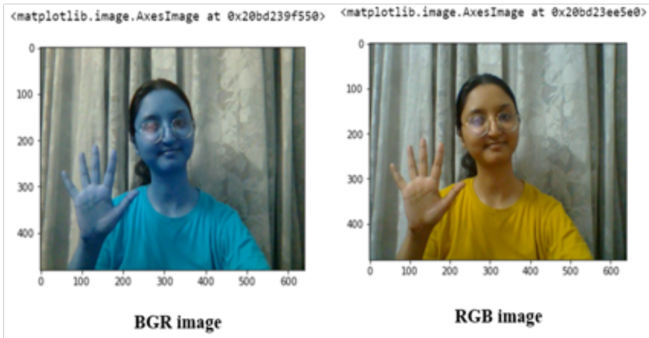


Figure 1 Conversion of images from BGR to RGB

Figure 2 Detection of key points through mediapipe

After obtaining, and extracting all the frames with their landmarks and key points, they were merged (face pose + hands landmark) and saved as NumPy arrays. This helped in labeling the signs and allowed the model to easily classify a gesture. Lastly, a loop was made, and all images were labeled according to their given signs. It should be noted that each sign has its own folder, which has thirty videos of 30 frames each. So, each video was labeled according to the sign; in total, every sign will have 900 frames.

Modeling

For modeling, an LSTM (LongShort-Term Memory) was used. LSTM is an advanced version of the RNN

(Recursive Neural Network) model, LSTM solves the main problem of RNN, which is the vanishing gradient problem [21]-[22]. In RNN, the gradient decreases while progressing through the network and it becomes more difficult to train the weights, meaning in later stages, the model does not learn much and cannot progress further. LSTM solves the forementioned problem by only remembering the important information learned from the previous run or sequence [23]. Another reason for choosing this particular model is its large number of parameters, so there is less need for fine adjustments. Furthermore, the first splitting of training and testing data was done into a split of 50%. Before training the LSTM model, the author logged into the tensor board dashboard to record the loss and the best seed value as the model trains over an epoch value of 800. An epoch is a phrase used in machine learning that denotes the number of runs the machine learning model has made across the full training dataset [24]. These are also referred to as iterations.

Further, as shown in Figure 3 LSTM model was applied and MSL was recognized in real-time.

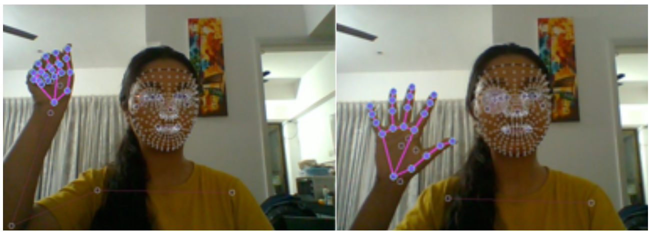


Figure 3 Malaysian sign language recognition with LSTM

The model recognizes most of the signs accurately; for sentence prediction, a loop was created so that the model can join consecutive words to make different sentences. Though, it should be noted that a proper wait time is to be allocated towards each sign so that the model can accurately predict the sentence.

Evaluation

This phase requires assessing the software or the model built against its original business goals. The evaluation phase also involves assessing any other data mining results that have been generated [25]. Evaluating the model using various metrics will be carried out. Accuracy is generally the first measure that comes to mind. It's a straightforward measure that calculates the proportion of correct to incorrect predictions. Another metric that can be used is recall which gives the fraction is correctly identified as positive out of all positives [26]. Both can be calculated using the Confusion matrix, for example.

A higher number of true negatives and true positives in a confusion matrix implies that the model is accurately predicting the class values [27]. It can also be a good estimation of whether the model is overfitted or not. For results, all values of the confusion matrix lie in the true negative and true positive parts, meaning the model is predicting the values quite accurately as shown in Figure 4. But in this case, a very small group of signs is used, as the dataset becomes bigger and more signs are included, it cannot be said that the model will remain this accurate.

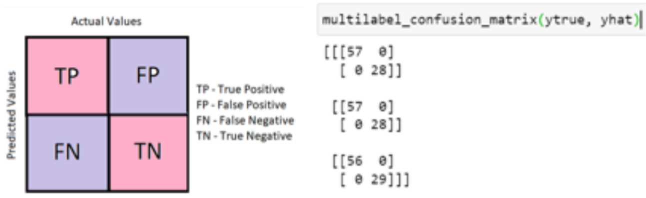


Figure 4 Confusion matrix for the MSL recognition results

The accuracy score, precision score and recall score of the model are very high for the same reason stated as shown in Figure 5.

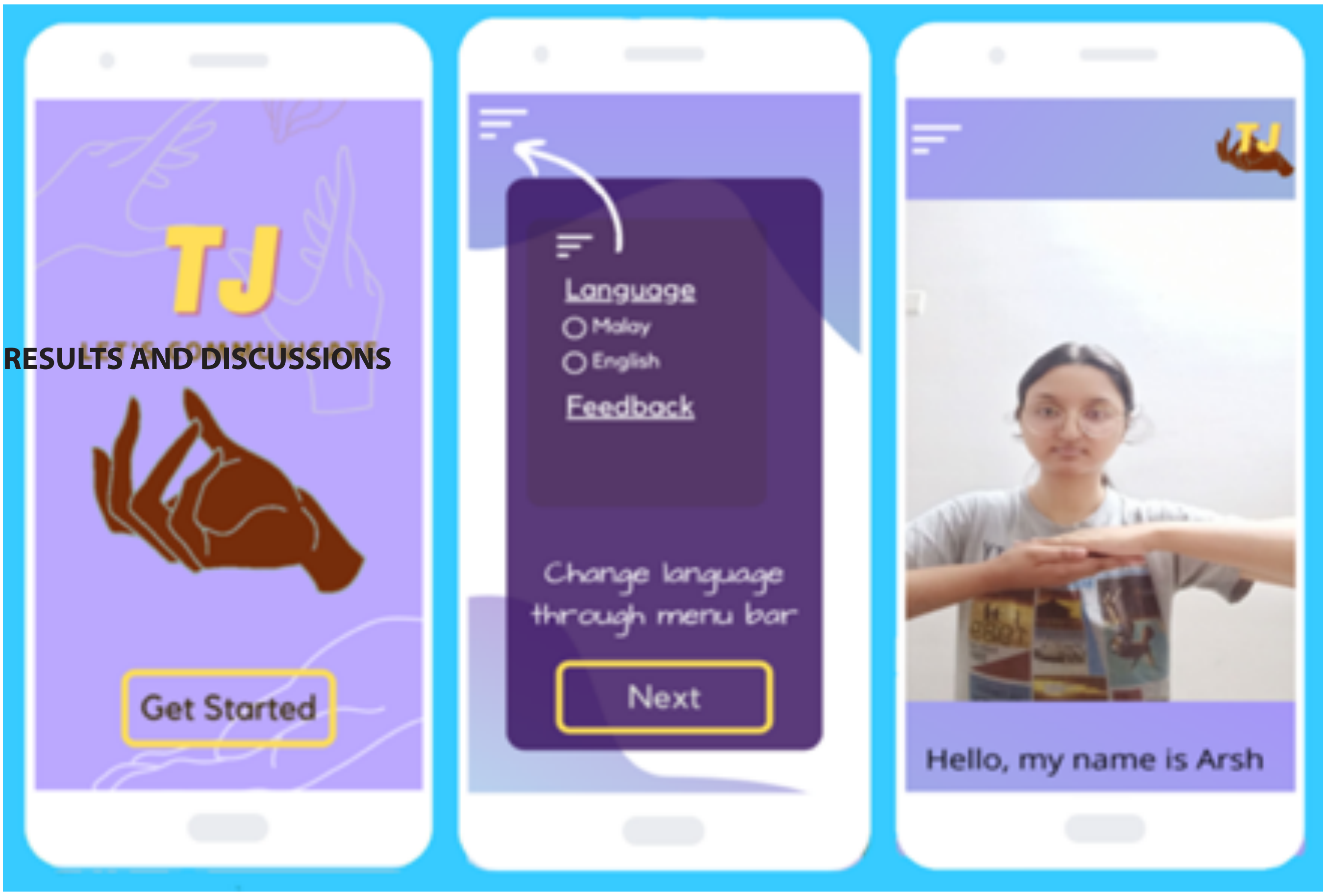
```
print("accuracy score is:" , accuracy_score(ytrue, yhat))
print("precision score is:" ,precision_score(ytrue, yhat,average='weighted'))
print("recall score is:" ,recall_score(ytrue, yhat, average='weighted'))

accuracy score is: 1.0
precision score is: 1.0
recall score is: 1.0
```

Figure 5 Accuracy, precision and recall scores for the model

Deployment Phase

After an iterative process of data preparation and modeling which leads to evaluation, the application would be deployed for the use of the target market. Monitoring and maintenance become critical if the data mining results become part of the day-to-day company and its surroundings. Preparing a maintenance strategy in advance helps to avoid excessively lengthy periods of inaccurate results. The best way to deploy the completed study would be Model deployment as a service; if the time and the resources are present, then the deployment is certainly possible. Figure 6 shows the proposed design for the application.



RESULTS AND DISCUSSIONS

Figure 6 Proposed design for Tuturjom application

RESULTS AND DISCUSSIONS

In this study, a demonstration of how hand and gesture detection can be achieved through machine learning was given. The media pipe holistic was used to extract and display various key points and landmarks that can be used in action recognition. Then LSTM was used to train the model, giving us quite a high accuracy considering the dataset provided. Figure 7 shown was obtained through a tensor board when the model was training. Here, the categorical accuracy of the model is shown as the epoch value changes over time. It can be seen that the accuracy of the model gradually increases and comes to a peak value of almost 100% at around 500. From there, the values are constant, except when they take a dip of 10% at the value of 600 and again increases to reach the maximum value. Here, the model is not overfitted, so an epoch value of 800 is considered good.

Figure 8 shows the loss experienced as the model is trained, it can be seen that the loss is quite high in the beginning but as the model trains and reaches the value of 800, the loss becomes almost negligent.

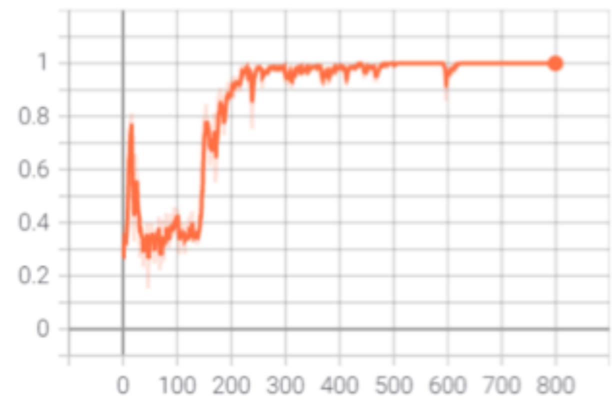


Figure 7 Epoch categorical accuracy for the model

It was also noted that the model had high precision and recall value which means there were less false positive and false negatives values.

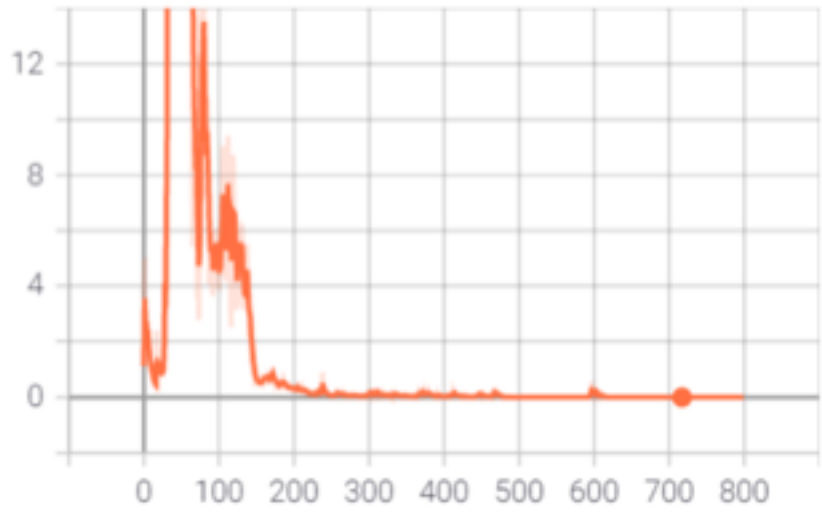


Figure 8 Epoch categorical loss for the model

LIMITATION OF THE STUDY

The application would be most efficient if a database containing proper pictures of MSL words and expressions existed. These images should be very clear, which becomes a limitation as MSL is currently not extensively researched, making it difficult to include such a database. It should also be noted that the model currently contains the basic words, letters and numbers for Malaysian sign language translation. As this database increases and more words are added, it can be said that the model parameters would have to be hyper-tuned accordingly.

The testing for this study was very limited; most users who tested the models were adults. Since the model is used for sign language determination and educating people about sign language. It is better to include children for user testing; this is to determine whether the model is able to detect the key points as accurately in the case of children as it is with adults.

CONCLUSION

The current situation necessitates the use of a machine-based sign language interpreter. People are indeed interested in learning Sign language, but not many sources are available. The problem can be solved by creating the real-time translation model. As a result, the suggested gesture recognition approach may pave the way for the creation of effective automated sign language translation systems. For those unable to speak or have hearing loss, such systems offer the potential to facilitate efficient communication and human–machine interaction. The most common areas where this may be employed are public venues such as ticketing desks, hospitals, etc. This may even be used to educate regular people in sign language.

The development of the current prototype offers an application which can translate MSL in real time. In future, there are hopes to include the feature of directly translated MSL text to speech. This will not only allow the application to work faster, but it will be more convenient for the user.

The author hopes to continue improving the dataset for a more precise translation and classification because having clean data will ultimately increase overall productivity and allow for the highest quality information in a model's decision-making. Also, efforts will be made to integrate the software into smaller hardware such as a wristwatch so that it is easier to use anywhere, and this will make communication through the application seamless.

REFERENCES

[1] A. Kumar, K. Thankachan, and M. M. Dominic, “Sign language recognition”, *3rd International Conference on Recent Advances in Information Technology (RAIT)*, 2016, doi: 10.1109/rait.2016.7507939

[2] R. Harini, R. Janani, S. Keerthana, S. Madhubala, and S. Venkatasubramanian, "Sign Language Translation", in *2020 6th International Conference on Advanced*

Computing and Communication Systems (ICACCS), 2020, doi: 10.1109/icaccs48705.2020.9074370

[3] A. Z. Shukor, M. F. Miskon, M. H. Jamaluddin, F. B. Ali@Ibrahim, M. F. Asyraf, and M. B. B. Bahar, A New Data Glove Approach for Malaysian Sign Language Detection", *Procedia Computer Science*, vol. 76, pp. 60–67, 2015, doi: 10.1016/j.procs.2015.12.276

[4] H. Hafit, C. W. Xiang, M. M. Yusof, N. Wahid, and S. Kassim, "Malaysian Sign Language Mobile Learning Application : A recommendation app to communicate with hearing-impaired communities", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, p. 5512, 2019, doi: 10.11591/ijece.v9i6.pp5512-5518

[5] H. A. Q. Maarif, R. Akmeliawati, and S. Bilal, "Malaysian Sign Language database for research", *International Conference on Computer and Communication Engineering (ICCCE)*, 2012, doi: 1109/iccce.2012.6271327

[6] F. Wong, G. Sainarayanan, W. M. Abdullah, A. Chekima, F. E. Jupirin, and Y. F. A. Gaus, "Software-Based Malaysian Sign Language Recognition", *Advances in Intelligent Systems and Computing*, pp. 297–306, 2013, doi: 10.1007/978-3-642-32063-7_31

[7] J. J. Ojeda-Castelo, M. D. L. M. Capobianco-Uriarte, J. A. Piedra-Fernandez, and R. Ayala, "A Survey on Intelligent Gesture Recognition Techniques", *IEEE Access*, vol. 10, pp. 87135–87156, 2022, doi: 10.1109/access.2022.3199358

[8] Z. H. Chen, J. T. Kim, J. Liang, J. Zhang, and Y. B. Yuan, "Real-Time Hand Gesture Recognition Using Finger Segmentation", *The Scientific World Journal*, pp. 1–9, 2014, doi: 10.1155/2014/267872

[9] J. P. Sahoo, A. J. Prakash, P. Pławiak, and S. Samantray, "Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network", *Sensors*, vol. 22, no. 3, p. 706, 2022, doi: 10.3390/s22030706

[10] T. Chouhan, A. Panse, A. K. Voona, and S. M. Sameer, "Smart glove with gesture recognition ability for the hearing and speech impaired", *IEEE Global Humanitarian Technology Conference – South Asia Satellite (GHTC-SAS)*, 2014, doi:10.1109/ghtc-sas.2014.6967567

[11] S. Pramada, Intelligent Sign Language Recognition Using Image Processing", *IOSR Journal of Engineering*, vol. 03, no. 02, pp. 45–51, 2013, doi: /10.9790/3021-03224551

[12] S. Khan, M. E. Ali, S. Das, and M. M. Rahman, "Real Time Hand Gesture Recognition by Skin Color Detection for American Sign Language", *4th International Conference on Electrical Information and Communication Technology (EICT)*, 2019, doi: 10.1109/eict48899.2019.9068809

[13] M. Al-Faris, J. Chiverton, D. Ndzi, and A. I. Ahmed, "A Review on Computer Vision-Based Methods for Human Action Recognition", *Journal of Imaging*, vol. 6, no. 6, p. 46, 2020, doi: 10.3390/jimaging6060046

[14] F. Gers, J. Schmidhuber, F. Cummins, "Learning to Forget: Continual Prediction with LSTM", *Neural Computation*, vol. 12, no. 10, 2000.

[15] N. L. Hakim, T. K. Shih, S. P. Kasthuri Arachchi, W. Aditya, Y. C. Chen, and C. Y. Lin, "Dynamic Hand Gesture Recognition Using 3DCNN and LSTM with FSM Context-Aware Model", *Sensors*, vol. 19, no. 24, p. 5429, 2019, doi:10.3390/s19245429

[16] S. Ameur, A. Ben Khalifa, and M. S. Bouhlel, "A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with Leap Motion", *Entertainment Computing*, vol. 35, p. 100373, 2020, doi: 10.1016/j.entcom.2020.100373

[17] A. Ikram, and Y. Liu, "Skeleton Based Dynamic Hand Gesture Recognition using LSTM and CNN", *2nd International Conference on Image Processing and Machine Vision*, 2020, doi: https://doi.org/10.1145/3421558.3421568

[18] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model", *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, doi:.1016/j.procs.2021.01.199

[19] F. Martinez-Plumed, L. Contreras-Ochando, C. Ferri, J. Hernandez-Orallo, M. Kull, N. Lachiche, M. J. Ramirez-Quintana, and P. Flach, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories", *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 8, pp. 3048–3061, 2021, doi:10.1109/tkde.2019.2962680

[20] Z. S. Abdallah, L. Du, and G. I. Webb, "Data Preparation", *Encyclopedia of Machine Learning and Data Mining*, pp. 318–327, 2017, doi:10.1007/978-1-4899-7687-1_62

[21] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model", *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, 2020, doi: 10.1007/s10462-020-09838-1

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[23] S. H. Noh, "Analysis of Gradient Vanishing of RNNs and Performance Comparison", *Information*, vol. 12, no. 11, pp. 442, 2021, doi: doi.org/10.3390/info12110442

[24] S. Afaq, and S. Rao, "Significance Of Epochs On Training A Neural Network", *International Journal of Scientific & Technology Research,* vol. 9, pp. 485-488, 2020.

[25] A. Nadali, E. N. Kakhky, and H. E. Nosratabadi, Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system", *3rd International Conference on Electronics Computer Technology*, 2011, doi:10.1109/icectech.2011.5942073

[26] H. Haibo and E. A. Garcia, Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/tkde.2008.239

[27] K. K. Al-jabery, T. Obafemi-Ajayi, G. R. Olbricht, and D. C. Wunsch II, "Data analysis and machine learning tools in MATLAB and *Python", Computational Learning Approaches to Data Analytics in Biomedical Applications*, pp. 231–290, 2020, doi: 10.1016/b978-0-12-814482-4.00009-7