Self-practice Khon Application using Artificial Intelligence

1st Thanadon Poomkosum
Suankularb Wittayalai Rangsit School
Pathum Thani, Thailand
17890@skr.ac.th

2rd Supachok Ruensom
Suankularb Wittayalai Rangsit School
Pathum Thani, Thailand
16743@skr.ac.th

3rd Nathamon Thiengcharoen
Suankularb Wittayalai Rangsit School
Pathum Thani, Thailand
16382@skr.ac.th

4th Nattapon Buaurai
Suankularb Wittayalai Rangsit School
Pathum Thani, Thailand
krunattapon@skr.ac.th

5th Anun Wongsang
Suankularb Wittayalai Rangsit School
Pathum Thani, Thailand
anun09042530@skr.ac.th

6th Somchoke Ruengittinun
Kasetsart University
Bangkok, Thailand
somchoke.r@gmail.com

***Abstract*—In Thailand, Khon institutions and professionals are limited. It is complex to practice with yourself; always practicing and getting recommendations is necessary. Utilizing artificial intelligence to construct a virtual expert to assist in practicing Khon dance with yourself more accurately was another possibility for solving problems. There were five stages in the work procedure. 1) Data collection involved gathering a Khon dance video of 25 postures, with 30 videos in each posture, from the internet and taking videos. 2) Extract frames from all videos. Then, divide into three datasets for 3 AI models based on the types of Khon dance postures: model 1 for basic movements, model 2 for dance vocabulary, and model 3 for sign language. Each dataset was divided into training (80%) and testing (20%) categories. 3) Train AI models; Long Short-Term Memory (LSTM) was used in AI model layer. 4) Evaluate models using accuracy and F1-score. 5) Calculate the important angles of each important movement in posture.**

We have developed a web application with a video and article page for learning how to dance, a training page for recording user videos to test dance movements, and a practice results page to view performance history. The program processes the video clips and then shows areas for improvement in each step and an accuracy score as a percentage. The application was evaluated by testing with a sample group of students from Suankularb Wittayalai Rangsit School, which had insufficient Khon classes. The performance of models indicated mean loss during training, Accuracy, and F1-score. The results have mean loss during training (mean loss during training: model1, model2, model3) (0.00931, 0.11687, 0.69274), accuracy (Accuracy: model1, model2, model3) (0.819896, 0.761989, 0.709120), and F1-scored (F1-score: model1, model2, model3) (0.821479, 0.813097, 0.700908). From user testing, the most accurate dance poses are Tob Khao, which is 82.03%; second, Iang Srisa, which is 79.39%; and third, Angry (Monkey), which is 78.30%.

Keywords—Khon, Artificial Intelligence, Deep Learning, Pose Estimation, MediaPipe

I. Introduction

Khon dance is a Thai classical art. Thai and foreigners are interested in Khon dance, but Khon institutions and professionals are limited in Thailand. Practicing alone can be challenging due to the complexity of always practicing and getting recommendations. Include those who had learned Khon already but had limited time. When you return to your instructor, a certain amount of time has passed, leading to less advancement in your training and a higher risk of injury if you train incorrectly.

For example, Suankularb Wittayalai Rangsit School students have a basic art class that combines theory and practice. Each class is only 50 minutes long. Due to the limited time for practice, students tend to learn only theory but lack sufficient skills for actual movements. In many cases, some schools face a shortage of Khon teachers, making it difficult for students to start themselves in Khon. In addition, private Khon dance lessons are expensive. Therefore, we have developed an AI-based self-learning application that can help these trainees who cannot manage their movements after attending classes at school. By accessing the system via the website anytime, anywhere, they can participate in their practice without the supervision or assistance of others. This software acts as a "virtual advisor" that helps students or users recognize their mistakes and judge their movements' accuracy.

II. Literature Review

A. *Khon dance*

Khon is an ancient Thai high-class dance that dates back to the Ayutthaya era. According to evidence from the memoirs of La Loubère, the French ambassador during the reign of King Narai the Great, Khon is a dance performed with movements to the sound of the saw and other musical instruments. The dancers wear masks and hold weapons. The important characteristic of Khon is that all the performers must wear masks, except for the male, female, and angel characters. The performance is only of the Ramakien story. [3]. There is no specific number of postures. So, we consulted with experts in Khon and selected 25 postures that can be practiced by oneself and are not dangerous during practice, divided into three categories. The first basic movements are Tob Khao, Ten Sao, and Tong Sa-ew. Second dance vocabulary, which are Kradok Thao, Krathung Thao, Kao Na, Jarod Thao, Tae Thao, Yok Thao, Jeep Kwam, Jeep Prok Khang, Jeep Prok Na, Jeep Song Lang, Jeep Ngai, Kod Lai, Lak Khor, and Iang Srisa. Third sign language, which are hate (ladies), hate (lords), hate (ogres), hate (monkeys), angry (ladies), angry (lords), angry (ogres), and angry (monkeys).

Fig. 1. Khon performs. [3].

B. Mediapipe

MediaPipe is an open-source platform developed by Google that is designed for research and development in multimodal data processing, including the analysis of images and videos. This platform uses deep learning to input data from images or videos, transforming them into body joint coordinates shown as x, y, and z values. [4]. In this project, Pose landmark detection and Hand landmark detection are utilized to extract body joint coordinates from videos of the Khon dance, as shown in Fig. 2. and Fig. 3.

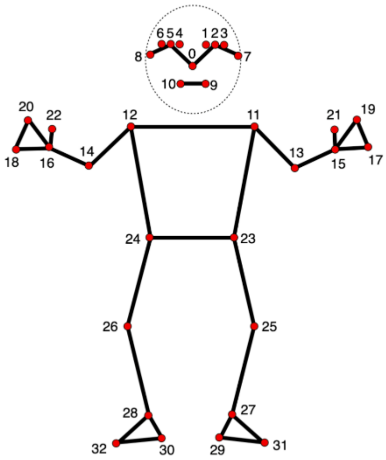

Fig. 2. Mediapipe's Pose landmark detection nodes. [6].

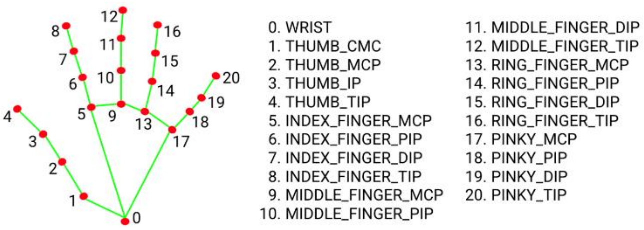

Fig. 3. Mediapipe's Pose landmark detection nodes. [5].

C. Long short-term memory

LSTM is an improvement of traditional Recursive Neural Networks (RNN), which has a problem of learning long-term dependency. LSTM solves this problem by adding long-term memory cells, which can store data for extended periods. The structure of LSTM has three gates; the First gate is called the forget gate, which removes unnecessary data from memory cells; the Input gate adds data to memory cells; and the output gate extracts valuable data from the current memory cell to output and uses as input for the next cell. [1]. So, It made LSTM learn long-term dependency and time-series data. Fig.4 shows the LSTM Network structure.

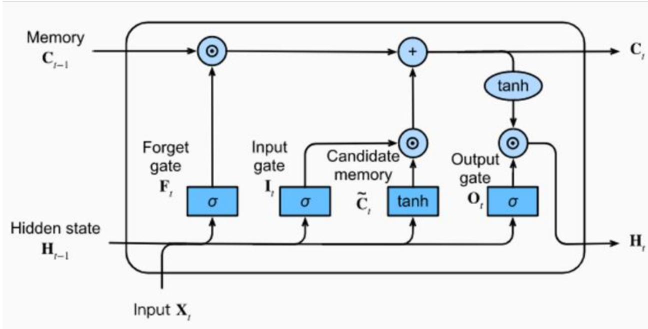

Fig. 4. Structure of an LSTM network. [2]

D. Calculate the angle in 3D space

The process of calculating angles in 3D space for pose estimation utilizes vectors in Mathematics. Start by assigning 3 points in 3D space and building vectors from these points. Then, the dot product of the vector is calculated, and the angle from the inverted cosine function is calculated. The result can be expressed in either radians or degrees. This method allows for highly accurate angle measurements. [7]. Fig.5 shows an angle between a pair of lines in 3D space.

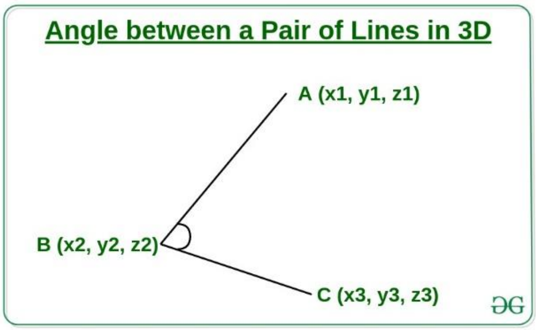

Fig. 5. An angle between a pair of lines in 3D. [7].

III. METHODOLOGY

To develop the application for practicing Khon, we need data to train the model and identify the similarities in Khon's performances. First, we must create a dataset using videos of 25 Khon poses, as previously indicated. Then, we find the joint angles, calculate the angles for all poses, and train the model with this dataset to evaluate its performance. After that, we will design the UX, UI and test the application to identify improvement and development, as shown in Fig. 6.

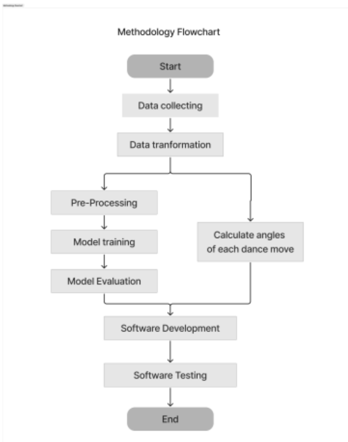

Fig. 6. Methodology's flowchart.

A. Data collecting

1) Dataset: we gathered 25 Khon dance poses online and filmed some poses. Each pose has 30 videos collected. An expert in Khon evaluated these videos. Once the experts assessed the correctness of the gathered information, the collected videos were categorized into 26 classes: 25 classes based on the specific Khon poses and 1 class for those that were not similar to any pose, including those that were performed incorrectly.

2) Data Transformation: From the previous, separate data in each class. Then, extract the frame by setting the delay, which captures the frame every 20 milliseconds. MediaPipe processes this image to obtain virtual joint values for each frame. Any joint value missing during a specific period is replaced with 0 in the empty value. Each frame's obtained virtual joint nodes are saved into a NumPy array file.

B. Pre-processing

1) Take all the data from each pose and select 30 frames from the total recorded frames using the frame-skipping technique.

2) Dividing it into three sets to train three models, where Model 1 is for basic movements, Model 2 is for dance vocabulary, and Model 3 is for sign language.

3) Split the data for each class to train and test the model, allocating 80% of the data for training and 20% for testing.

C. Models training

To train each model, the model layer is set as in Fig. 7, and the epoch value is set to 1500.

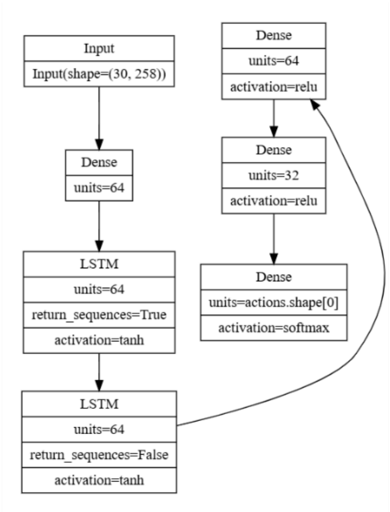

Fig. 7. The AI model layer.

D. Models evaluation

To evaluate the model, we use Accuracy to assess how well the model performs in classifying with the same amount of data in each class and F1-score to assess how well the model performs in classifying in each class, according to Equations (1), (2), (3), and (4). [8].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Reall} \quad (4)$$

E. Calculating the Angles of Each Dance Move

1) Feature extraction: Take all the data of each pose and select the frames that are the important moments of that pose. Then, from the selected data, calculate the angles of body nodes in 3D space using equations (5), (6), (7), and (8) [7] by selecting the important angles of that moment.

$$dot = (x_1 - x_2) \cdot (x_3 - x_2) + (y_1 - y_2) \cdot (y_3 - y_2) + (z_1 - z_2) \cdot (z_3 - z_2) \quad (5)$$

$$mag_A = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (6)$$

$$mag_B = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2 + (z_3 - z_2)^2} \quad (7)$$

$$\theta = \cos^{-1}\left(\frac{dot}{mag_A \cdot mag_B}\right) \cdot \frac{180}{\pi} \quad (8)$$

2) Angle comparison: We choose the angle with the greatest or least value at the moment. For example, in the moment of Tob Khao with the right hand raised, we will compare the greatest degree of the right shoulder angle to make it easier to advise for each moment of the user. If there is a deviation from the standard of more than 10 degrees, the advice status will be "must improvement." If there is a deviation from the standard of more than 5 degrees but not more than 10 degrees, the advice status will be "should improvement." .

F. Software development

1) Tools for development software

a) We use HTML, CSS, and JavaScript to develop front-end web applications, along with the Mediapipe.js library to receive virtual connector values from the client side to send smaller data to the server.

b) We use Python with the FastAPI web framework for the backend to communicate and handle user requests to the database via RestAPIs and the TensorFlow library for AI-powered data processing.

c) We use the MySQL database to save data and SQL language to manage data in the database.

2) Web application workflow design: first, users log in to the web application. The system shows a list of all Khon postures. After users choose a posture to practice, the system will show a video tutorial to study the correct way to perform it. Next, users will go to the training page and record a video of their dance. After that, the system will process the data and display the similarity value as a percentage. If the user wants to view their training history or see their posture improvement suggestions, go to the training results page and then select their training round, as shown in Fig. 8 - 14.

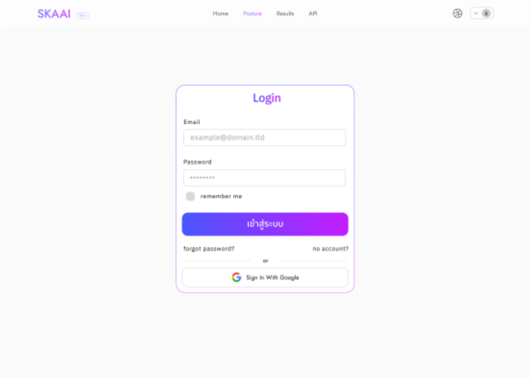

Fig. 8. The login page.

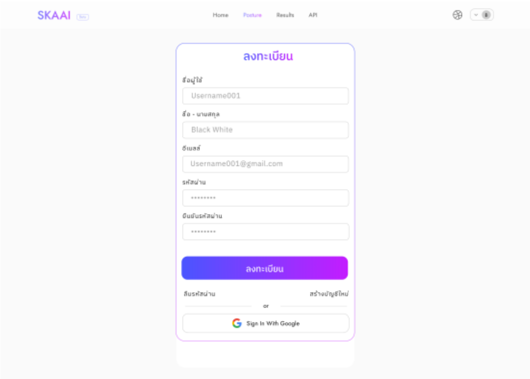

Fig. 9. The registration page.

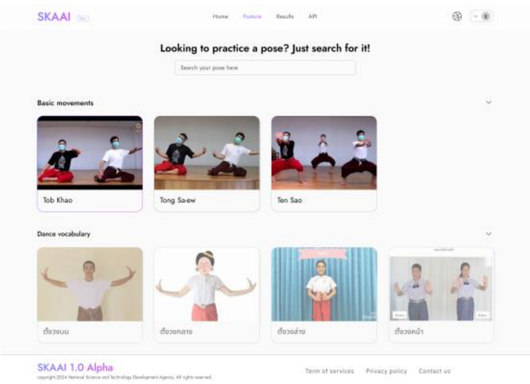

Fig. 10. The main page shows you can choose the posture to practice.

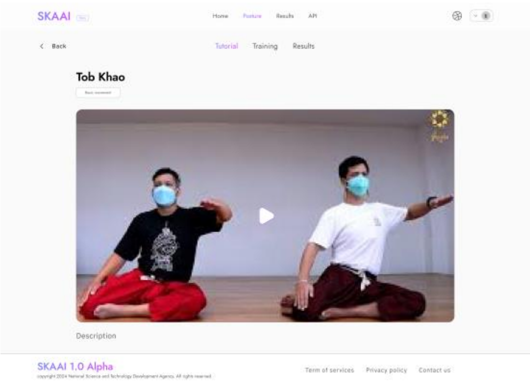

Fig. 11. The tutorial page.

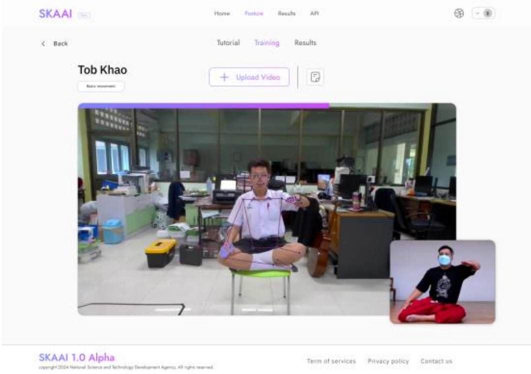

Fig. 12. The training page during training.

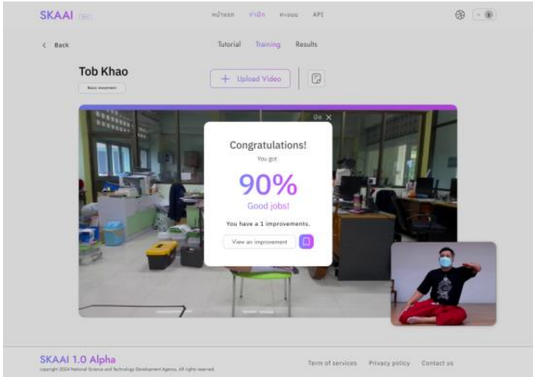

Fig. 13. The results show a similarity score after training.

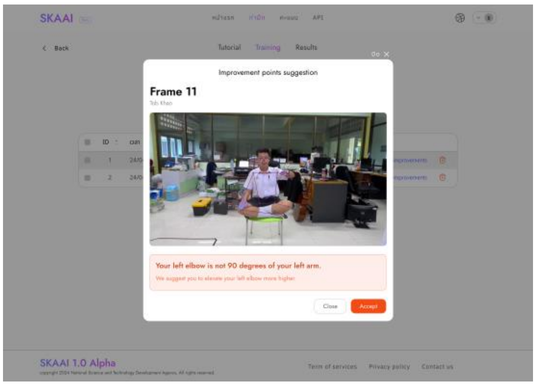

Fig. 14. The results analysis page that suggests improvement points.

G. Software testing

It was tested from a sample group of 21 students from the Khon Harnsa Club of Suankularb Wittayalai Rangsit School and 24 students from the Natasilphaploen Club of Suankularb Wittayalai Rangsit School, totaling 45 people, divided into 30 people who had previously trained in Khon dancing and 15 people who were beginning to train. Data was collected on the average accuracy of users who practiced each Khon dance move three times. The average was found when the Khon expert assessed that the user had done everything correctly.

IV. RESULT

The application's performance was divided into three parts: 1) evaluating the AI model indicated mean loss during training, accuracy, and F1-score of all three models. It was found that Model 1 is the most efficient, and the average accuracy of all three models is 0.7600, or 76.00%, as shown in Table I, which shows the AI model performance in each model.

TABLE I. PERFORMANCE OF EACH MODEL

Model	Mean loss during training	Accuracy	F1 score
Model 1	0.00931	0.819896	0.821479
Model 2	0.11687	0.751109	0.761989
Model 3	0.69274	0.709120	0.700908

2) The sample group results collected the average accuracy data of the users who practiced the Khon dance moves. When the Khon experts assessed that the users did everything correctly, they found that the average accuracy collected from the entire sample group was 72.3784%, as shown in Table II.

TABLE II. MEAN ACCURACY IN EACH POSTURE

posture	mean Accuracy (Already trained) [%]	mean Accuracy (Never trained) [%]	Mean Accuracy [%]
1. Tob Khao	84.93	79.13	82.03
2. Ten Sao	80.50	61.89	71.20
3. Thong Sa-eo	73.87	67.34	70.61
4. Kradok Thao	82.40	72.36	77.38
5. Krathung Thao	68.27	60.05	64.16
6. Kao Na	75.07	72.37	73.72
7. Jarod Thao	69.40	62.00	65.70
8. Tae Thao	75.73	71.26	73.50
9. Yok Thao	73.77	70.60	72.19
10. Jeep Kwam	71.80	69.96	70.88
11. Jeep Prok Khang	74.05	72.92	73.49
12. Jeep Prok Na	70.90	76.47	73.69
13. Jeep Song Lang	70.87	74.57	72.72
14. Jeep Ngai	70.23	70.87	70.55
15. Kod Lai	73.56	73.50	73.53
16. Lak Khor	75.56	75.56	75.56
17. Iang Srisa	78.28	80.49	79.39
18. Hate (lady)	72.05	65.37	68.71
19. Hate (lords)	75.27	60.57	67.92
20. Hate (ogres)	72.99	66.60	69.80
21. Hate (monkeys)	72.40	65.03	68.72
22. Angry (ladies)	77.37	70.25	73.81
23. Angry (lords)	72.99	68.75	70.87
24. Angry (ogres)	72.09	69.54	70.82
25. Angry (monkeys)	80.73	75.87	78.30

3) The average score from PageSpeed Insights evaluation in each category is 94.19 to 100 points, which means all web pages are at a reasonable level, with M referring to mobile and D referring to desktop, as shown in Table III.

TABLE III. PAGESPEED INSIGHT PERFORMANCE TESTING

Web Pages	Performance		Accessibility		Best Practices		SEO	
	M	D	M	D	M	D	M	D
Login	90	93	100	100	100	100	100	100
Register	95	95	100	100	100	100	100	100
Exercises	87	91	100	100	96	96	100	100
Exercise tutorial	90	91	91	91	100	100	100	100
Exercise training	97	98	92	92	96	96	100	100
Exercise results	93	99	100	91	96	96	100	100
Setting	97	99	97	99	97	98	100	100
Contacts	96	93	100	100	100	100	100	100

V. CONCLUSION

From the testing results of the mean loss during training, Accuracy, and F1-score of the three models, it can be concluded that the model with the best test results is model 1 with a mean loss during training of 0.00931, Accuracy of 0.819896, and F1-score of 0.8221479, followed by model 3 with a mean loss during training of 0.69274, accuracy of 0.709120, and F1-score of 0.700908, and finally model 2 with a mean loss during training of 0.11687, accuracy of 0.751109, and F1-score of 0.761989. From the test results with a sample group of students from Suankularb Wittayalai Rangsit

School, the three most effective postures are the Tob Khao posture, accounting for 82.03%, followed by the Iang Srisa posture, accounting for 79.39%, and the angry pose (monkey) is 78.30%. The results of the web application performance testing via the PageSpeed Insights tool found that all webpages scored 90-97 in the performance category, 91-100 in the accessibility category, 91-100 in the best practices category, and 100 in the SEO category for all webpages. The results of the PageSpeed Insights test show that the project has excellent performance and accessibility in all categories, including performance, accessibility, best practices, and SEO, all of which align with best practices and international standards for web development. This ensures that the webpages in this project will provide a good user experience in all aspects, support accessibility for all user groups, including those with limited access, and comply with security standards and best design for users on all devices.

VI. ACKNOWLEDGMENT

We would like to thank the Department of Computer Science at Kasetsart University for their support, suggestions, and meticulous correction of errors, as well as for providing knowledge on software development that ensured the project's completion. This project is part of an academic collaboration between Suankularb Wittayalai Rangsit School and the Department of Computer Science at Kasetsart University.

Our gratitude also goes to Suankularb Wittayalai Rangsit School and Information Communication and Technology Center, Suankularb Wittayalai Rangsit School, for their support, assistance, and opportunities for students to participate in the competition and submit articles.

We appreciate the National Science and Technology Development Agency's funding for the 26th National Software Contest for the Self-practice Khon Application using Artificial Intelligence project.

References

[1] Aakarshachug, "What is LSTM – Long Short Term Memory?," GeeksforGeeks, Jun. 10, 2024. [Online]. Available: https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/. [Accessed: Aug. 27, 2024].

[2] Calzone, O., "Structure of an LSTM Network," in An Intuitive Explanation of LSTM. Medium, Feb. 21, 2022. [Online]. Available: https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c. [Accessed: Sep. 1, 2024].

[3] Department of Cultural Promotion, "Khon," Intangible cultural heritage of Thailand, (n.d.). [Online]. Available: https://ich-thailand.org/heritage/detail/6291e5b7978f238e61f77da4. [Accessed: Aug. 27, 2024].

[4] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., and Grundmann, M., "MediaPipe: A Framework for Perceiving and Processing Reality," Google Research, 2019. [Online]. Available: https://static1.squarespace.com/static/5c3f69e1cc8fedbc039ea739/t/5e130ff310a69061a71cbd7c/1578307584840/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Workshop_2019.pdf. [Accessed: Aug. 27, 2024].

[5] "Mediapipe's Hand landmark detection nodes," in Pose landmark detection guide. Google AI for Developers, Aug. 22, 2024. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker?hl=th. [Accessed: Sep. 3, 2024].

[6] "Mediapipe's Pose landmark detection nodes," in Pose landmark detection guide. Google AI for Developers, Aug. 22, 2024. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker?hl=th. [Accessed: Sep. 3, 2024].

[7] Rituraj Jain, "Python | Implementing 3D Vectors using dunder methods," GeeksforGeeks, May. 31, 2021. [Online]. Available: https://www.geeksforgeeks.org/python-implementing-3d-vectors-using-dunder-methods/. [Accessed: Sep. 4, 2024].

[8] "Understanding the Confusion Matrix in Machine Learning," GeeksforGeeks, Jul. 8, 2024. [Online]. Available: https://www.geeksforgeeks.org/confusion-matrix-machine-learning/. [Accessed: Sep. 2, 2024]