



Motion Capture 3D Wayang Character with Pose Landmarks using MediaPipe

Arnanda Prasatya¹, Ferdi Ahmad Ariesta², Keanu Rayhan Harits³, Trisna Gelar^{4*}, Muhammad Rizqi Sholahuddin⁵, Iwan Awaludin⁶, Aprianti Nanda Sari⁷

¹⁻⁷Department of Computer and Informatics, Politeknik Negeri Bandung, Bandung, Indonesia

Correspondence: E-mail: trisna.gelar@polban.ac.id

ABSTRACT

Motion Capture (MoCap) is vital for the digital tracking and recording of motions across many applications, enhancing sectors from entertainment to industrial processes. This work developed a web-based markerless motion capture system utilizing MediaPipe, capable of recognizing and replicating real-time human movements. The movements are then displayed on a 3D Wayang character, Semar, rendered in WebGL. The process was to record live webcam footage, process it with MediaPipe to find landmarks for 33-point poses, project these 3D coordinates onto the Blender-rigged Wayang model, and then use WebGL to render the animated result. The results showed that accuracy was greatly affected by ambient lighting and that the system worked best between 0.5 and 3 meters from the camera. Although hand landmark detection produced accurate results, the intricacy of the rigging and the absence of specific WebGL 3D animation references made it difficult to accurately integrate leg and body movements with the 3D model. Regardless the challenges, this research provides a more practical and cost-effective alternative to conventional motion capture methods, opening up new possibilities for entertainment, gaming, and cultural heritage preservation. Character rigging will be improved and MediaPipe will be further integrated with other 3D models in the future.

ARTICLE INFO

Article History:

Submitted/Received

28 March 2025

First Revised 15 April 2025

Accepted 01 May 2025

First Available online

01 June 2025

Publication Date 01 June 2025

Keyword:

3D Pose Estimation;

Markerless;

MediaPipe;

Motion Capture;

Wayang Character.

1. INTRODUCTION

The Industrial Revolution 4.0 brought about a lot of new technologies and ideas for industrial processes, which means that we need more advanced ways to make them happen. Motion Capture (MoCap) is one of the most well-known uses of technology in this field. Motion capture employs several technologies, including RGB cameras, infrared cameras, depth cameras, and inertial measurement units (IMUs), to monitor and document the motions of individuals and objects in real time. Motion capture technology can enhance automated production, monitor worker productivity, and improve workplace safety (Menolotto et al., 2020).

Computer technology has significantly transformed the interaction between individuals and computers in this evolving environment. Three-dimensional (3D) Human Pose Estimation is a recognized computer vision method that utilizes images or videos to reconstruct the configurations of human skeletons in three-dimensional space. This approach enables the creation of virtual models of human skeletons, which is crucial for motion capture (MoCap). The primary objective of 3D pose estimation in motion capture is to record human body movements and convert them into a skeleton or 3D model applicable in several domains, including robotics, sports, entertainment, healthcare, and computer animation (Desmarais et al., 2021).

The most important thing about this 3D pose estimation technology in MoCap is that it lets you naturally recreate how people move in 2D images or videos without needing extra hardware like reflectors or special cameras. 3D pose estimation has become a major area of research in the last few years because it can be used in so many different fields (Lin et al., 2023). MediaPipe (Lugaresi et al., 2019), a framework for detecting human stances and postures in photos and videos, offers a more sophisticated method for capturing motion in videos. MediaPipe is highly advantageous since it can swiftly and accurately recognize body and user positions in real time, without requiring specialized sensors or markers.

This study will examine the application of MediaPipe-based motion capture technologies to identify and replicate human body movements onto 3D character models. The 3D model for this project will be Semar, a traditional Indonesian Wayang character recognized for its wisdom and guiding symbolism. The implementation utilizes MediaPipe's posture landmarks to identify and analyze the locations of human body joints in real time. MediaPipe can enhance the precision of 3D pose landmarks to over 90%, hence increasing its use for video-based motion capture (Lin et al., 2023; Potenzianni et al., 2018).

Utilizing these posture markers facilitates the recreation of human movement and enables the implementation of the generated results in 3D applications rendered swiftly with WebGL. For instance, it facilitates the creation of dynamic 3D animated characters, such as the Wayang character, which are governed by recorded data on human body movements. This can be useful in a lot of areas, such as making cartoons and video games for fun (Awaludin et al., 2024; Sharma et al., 2019), looking at posture for rehabilitation or physical training in healthcare (Kim et al., 2023; Salisu et al., 2023; Scott et al., 2022), and more.

This study seeks to enhance markerless motion capture techniques with MediaPipe technology. Specifically, it demonstrates the application of MediaPipe with conventional 3D Wayang characters rendered in WebGL. This technique is anticipated to be a superior and more adaptable alternative to conventional motion capture approaches, particularly for animating culturally significant characters. Ultimately, the findings of this study should yield valuable technological solutions applicable across several industries and facilitate innovative methods for preserving cultural heritage through digital animation.

2. METHODS

2.1. Experimental Approach

This study employed an experimental methodology with implementation phases to evaluate the efficacy of MediaPipe's Motion Capture technology in detecting human body movements and mapping posture markers onto 3D character models. The objective of this method was to demonstrate that MediaPipe is an effective architecture for markerless motion capture in videos.

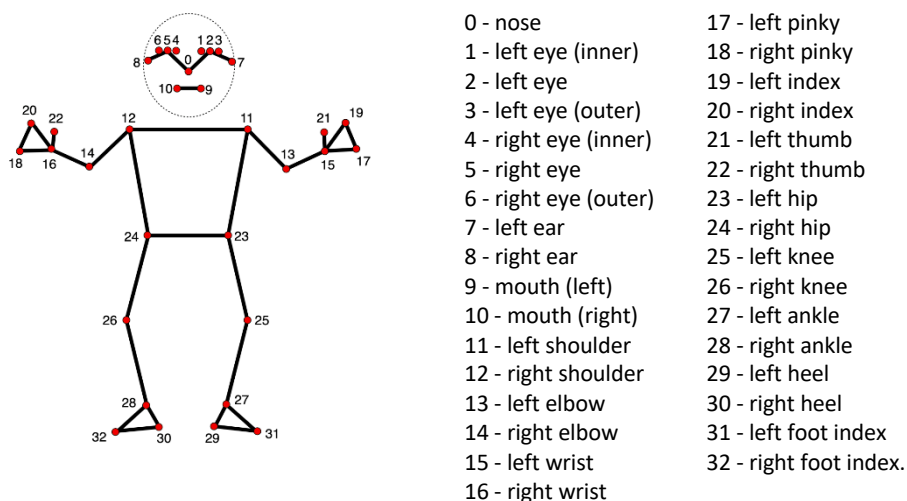


Figure 1. MediaPipe's 33 body landmark locations(Developers, 2025)

Figure 1 illustrates MediaPipe's 33 body landmark locations (Developers, 2025). MediaPipe Pose was utilized to detect and analyse 33 human body pose landmarks in real-time. These pose landmarks were mapped based on the joint positions identified by the MediaPipe algorithm. Subsequently, the detected landmark data was projected onto a 3D character model. This process involved mapping the pose landmarks' points to the skeleton structure of the 3D model using supporting software, namely Blender.

2.2. 3D Wayang Mocap Workflow

The workflow commenced with the activation of a webcam for real-time video acquisition. Subsequently, the system analysed these frames to identify crucial areas for pose estimation, with a focus on the face, body, and limbs. MediaPipe then detected these pose estimation areas and utilized its Holistic and Face Detection models to extract the corresponding landmarks. Following this, the extracted landmarks were meticulously integrated into an existing 3D character model through a rigging process. Finally, WebGL was employed to render the animated 3D object. This entire process was iteratively performed to ensure accurate and consistent motion reconstruction. Figure 2 depicts the complete workflow.

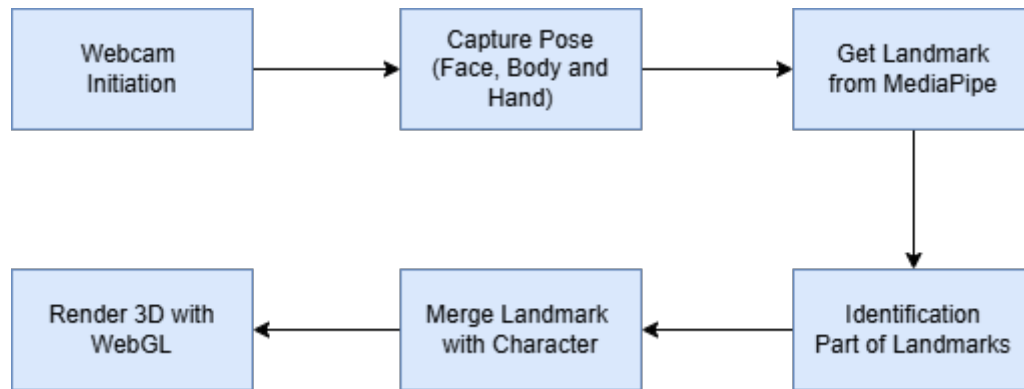


Figure 2. Workflow of 3D Wayng Mocap

2.3. Implementation 3D Wayang Mocap

The implementation of the motion capture system with MediaPipe involved four development stages:

1. Development Environment Setup

- JavaScript was chosen as the primary programming language for this project due to its extensive libraries, compatibility with web applications, and seamless integration with devices such as webcams. JavaScript's capabilities were well-suited for the real-time processing demands of a motion capture system.
- The project requires MediaPipe for posture landmark detection, OpenCV(Odeh & Odeh, 2024) for webcam video acquisition and image processing operations(Wiley & Lucas, 2018), and WebGL(Evans et al., 2014) for 3D pose model visualization in a browser. A camera provides rapid movement recording. OpenCV sources the video, which MediaPipe then processes frame by frame to extract pose information.

2. Implementation of Pose Detection

- MediaPipe has a Pose Solution that can find human pose landmarks in real time. This solution finds 33 points on the body, including 3D coordinates (x, y, z).
- Using OpenCV, a pipeline was set up to read frames from the webcam one at a time. MediaPipe then processed each frame to find pose landmarks.
- MediaPipe makes pose data by getting the 3D coordinates (x, y, z) of each landmark it finds. Next, these coordinates are used to make 3D models of movements.

3. 3D Model Development

- 3D Character Model Creation: The 3D character model utilized in this study is Semar, a traditional character from Indonesian Wayang. Semar is a prominent Javanese Wayang character symbolizing a guide or counselor for comprehending the essence of life. He is recognized for his wisdom and kindness(Gumilang & Robinson, 2022).
- Rigging Implementation on the 3D Character Model: After creating the 3D character, an armature bone human (meta-rig) was implemented for the 3D character. The selected armature bones were carefully chosen to align with

the pose detection shapes determined by the MediaPipe model. The rigging was then adjusted to closely resemble the character's form. Figure 3 illustrates the 3D Wayang Semar character.



Figure 3. 3D Wayang Semar character Model in Blender

- **Export Formatting:** Following the rigging process, the object file was exported in a compatible format for further processing, specifically glTF (.glb/.gltf) format. Both formats are compatible for subsequent processing involving armature bones.
4. System Integration
- Both the MediaPipe and WebGL libraries were developed using JavaScript. The integration process commenced with the configuration of MediaPipe Pose for landmark detection and WebGL for rendering 3D images. MediaPipe converted the camera video into 33 pose markers. These were later transformed into WebGL coordinates.
 - **MediaPipe Development for Landmark Transformation to 3D Model Space:** The system employed transformation matrices to convert the pose landmark coordinates from MediaPipe into the appropriate 3D spatial coordinates for the character model.

3. RESULTS AND DISCUSSION

This section discusses the findings of the study, including the development and evaluation of the web-based motion capture application. It emphasizes the capabilities of MediaPipe technology in detecting and reconstructing human body movements in three-dimensional space, subsequently elaborating on the outcomes and their implications.

3.1. Results

3.1.1 Pose Landmark Detection

This section details tests on Pose Landmark identification to demonstrate the system's proficiency in motion capture. Pose landmarks were identified by monitoring the user's entire body through free poses. Utilizing OpenCV to capture body movements via a webcam, subsequently processed by MediaPipe, the system effectively detected the user's movements

and combined them with the 3D object model. Below are instances of outcomes from pose landmark detection with free-position styles:

Table 1. Pose Detection Landmarks



Pose	Result
Free Pose	
Free Pose with leg lift	

Table 1 illustrates the system's capability to qualitatively identify diverse user pose movements, resulting from pose landmark detection. The results indicate that MediaPipe's posture landmarks successfully detected the user's leg motions; however, the integration with the 3D object model failed to track these movements, hindering the object's ability to follow the leg's motion.

3.1.2 Integration with 3D Model

Following the experiments on pose landmark detection with free-pose photographs, the next step involved integrating with the 3D model through trials on each body part. In this process, OpenCV, which reads the user's body movements from the webcam, was integrated with MediaPipe to detect pose landmarks, which subsequently animated the 3D object model.

The following are the qualitative results of the 3D model integration implementation, with experiments conducted on each body part.

Table 2. Pose Detection with 3D character

Pose	Result
Pose one hand	

Pose both hand



Pose body



3.2. Evaluation

Application's performance and reliability were qualitatively evaluated the in gesture detection and 3D model alignment by examining the following factors:


1. MediaPipe Detection Distance Testing

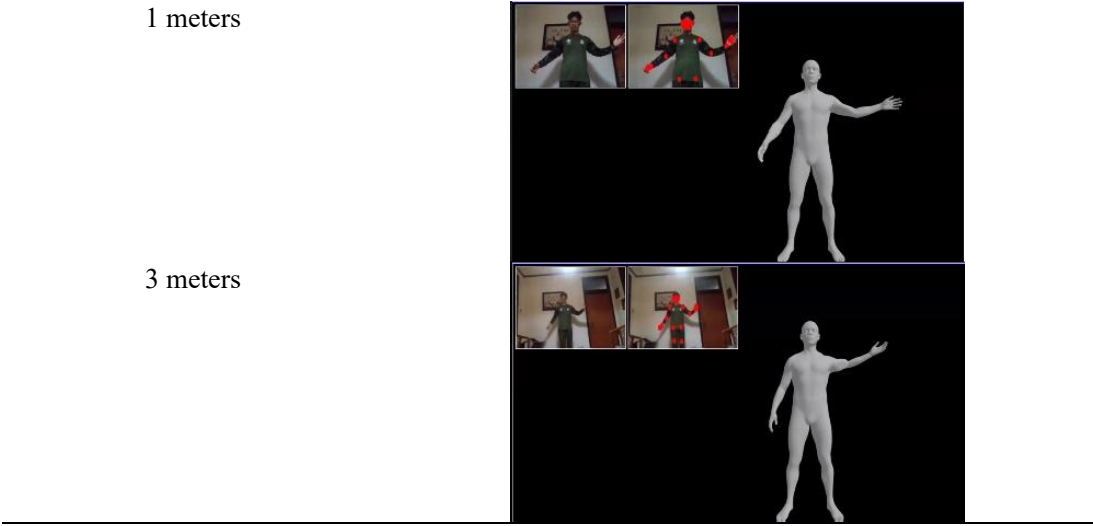
The objective of the initial test was to qualitatively evaluate the precision of body gesture detection. In this experiment, various usage scenarios were simulated across different distances, ranging from the closest distance (0.5 meters) to a far distance (3 meters).

Test Results:

- Close distance (0.5 meters): The application successfully detected body gestures with precise landmark placement.
- Ideal distance (1 meter): The application effectively identified body motions at an optimal usage posture, with MediaPipe's landmark representation being notably accurate.
- Far distance (3 meters): Even at a significant distance, the application was still able to detect landmarks with appropriate placement, as qualitatively evaluated.

Table 3. Evaluation Pose Detection with various distances

Distance	Result
0.5 meters	



2. Testing Environmental Lighting Conditions in MediaPipe Detection
The subsequent test involved qualitatively evaluating the application's performance in environments with suboptimal lighting conditions. This test aimed to ascertain the application's compatibility across various lighting environments.

Table 4. Evaluation Pose Detection with various light conditions

Lighting Conditions	Result
Bright and sufficient lighting conditions	
Adequate lighting condition	
Low light condition	

- Test Results:
- Bright and sufficient lighting conditions: In both bright and sufficiently lit environments, the application was able to detect pose estimation parts with precision.
 - Low light: On the other hand, when there was little light, the application showed strange pose estimates.

3. Correspondence of Detected Pose Movements with 3D Model




The final test qualitatively assessed the accuracy of the 3D model's correspondence with the provided pose estimation, specifically the alignment achieved through connecting landmarks with the 3D rigging. This test was performed with several poses. Test Results:

- In the two-hand movement test, the 3D character object could undergo changes, but the accuracy was not as precise as detected by MediaPipe.
- In the body movement and leg lift tests, there was no correspondence at all with the 3D character, indicating inaccurate body movements.

3.3. Discussion

The application of MediaPipe in this motion capture software for 3D character animation with position estimation revealed some significant insights. MediaPipe possesses advantageous characteristics, including a user-friendly library and extensive documentation that facilitate learning and utilization. The precision of posture landmark detection is significantly influenced by external factors, particularly the distance from the person to the camera and the ambient lighting conditions. These factors proved to be crucial for the system's capacity to precisely detect and monitor bodily movements.

Table 5. Evaluation Correspondence of Detected Pose Movements with 3D models.

Movements	Result
Body movement	
Both hand movement	
Leg lift movement	

The evaluation results indicated that the optimal distance for effective pose detection between the user and the camera ranged from 0.5 to 3 meters. The illumination of the setting was also highly significant. Illumination that is both bright and sufficient enhances detection

accuracy compared to poor lighting. Nonetheless, a significant disparity existed between the observed movements and the animation of the 3D model. This was particularly applicable to bodily and leg movements, which lacked precision. This was particularly applicable to bodily and leg movements, which lacked precision. This constraint primarily arose from several technical challenges, including:

- **Rotation Matrices Mismatch:** The direct translation of MediaPipe's 3D landmark coordinates into rotation matrices for the 3D model's bones proved challenging. Discrepancies between the coordinate systems or the inherent rotational properties of MediaPipe's output and the 3D model's armature could lead to misalignments.
- **Joint Hierarchy and Rigging Complexity:** The intricate joint hierarchy of the Semar character's rigging, especially for the legs and torso, may not have perfectly aligned with MediaPipe's simplified 33-point pose structure. This mismatch could result in unnatural or failed movements when attempting to drive complex joint rotations from a limited set of input points.
- **Absence of Inverse Kinematics (IK) Implementation:** The current system likely relies on forward kinematics (FK) or direct mapping. Without a robust Inverse Kinematics (IK) solver, achieving natural and constrained movements for limbs like legs (where the foot needs to stay grounded or follow a specific path) becomes significantly more difficult. IK would allow for more intuitive control over the end effectors (e.g., feet) and propagate movements up the joint chain, leading to more realistic animation.
- **Lack of Specific WebGL 3D Animation References:** As noted, the absence of explicit examples demonstrating advanced WebGL applications for 3D character animation, particularly concerning complex rigging and motion transfer, further complicated the precise development of this motion capture system. This necessitated a more foundational approach to mapping, which may not have fully addressed the nuances of realistic limb movement.

The MediaPipe technology is more user-friendly than alternative markerless motion capture methods such as iPi Soft, which employs RGB video sequences and deep convolutional neural networks. MediaPipe offers a superior and more cost-effective solution for markerless tracking using a standard webcam, whereas iPi Soft requires a more complex configuration involving two Kinect V2 sensors and specialized software (Min et al., 2019).

4. CONCLUSION

This study successfully developed a web-based markerless motion capture system utilizing MediaPipe to detect and recreate 3D human body movements. The system performed optimally within a distance range of 0.5 to 3 meters from the camera, and its accuracy was significantly influenced by ambient lighting conditions. While hand landmark identification functioned effectively, challenges arose in accurately recognizing and integrating leg and body movements with the 3D character rigging. These difficulties can be attributed to technical complexities such as potential rotation matrices mismatch, the intricate joint hierarchy of the 3D model, and the absence of a dedicated inverse kinematics (IK) solution for more natural limb control. The experimental results presented are based on qualitative observations, providing an initial assessment of the system's capabilities. Despite these challenges, the study enhances the usability and utility of motion capture technologies by offering a more practical and cost-effective alternative to conventional methods. Future work should focus on improving character rigging, looking into different types of 3D characters (like cartoons and Wayang), and doing more research on how to use MediaPipe-3D models in more areas

of entertainment, gaming, and cultural preservation. To provide a more rigorous and comprehensive evaluation, future research should incorporate quantitative metrics such as FPS, Latency, Accuracy (compared to ground truth) and Robustness under varying conditions. Conducting such quantitative analyses will enable direct comparisons with other markerless motion capture systems and provide a more objective assessment of the system's efficacy, further advancing its application in entertainment, gaming, and cultural preservation.

5. ACKNOWLEDGMENT

This research was supported by DIPA funds from the Bandung State Polytechnic, under the Applied Research funding agreement No. 235.4/PL1.R7/PG.00.03/2024. The authors extend their sincere gratitude to the dedicated students who assisted in the data collection and system implementation phases of this research.

6. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

7. REFERENCES

- Awaludin, I., Sholahuddin, M. R., Gelar, T., Engineering, I., and Bandung, P. N. (2024). 3D Large-Screen Immersive Video Mapping Installations & Interactive Light Game for Conference Seminars. *Journal of Computer Engineering , Electronics and Information Technology (COELITE)*, 3(2), 89–100.
- Desmarais, Y., Mottet, D., Slangen, P., and Montesinos, P. (2021). A review of 3D human pose estimation algorithms for markerless motion capture. *Computer Vision and Image Understanding*, 212, 103275.
- Developers, G. (2025). *Mediapipe 33 point Landmarks*. https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker
- Evans, A., Romeo, M., Bahrehmand, A., Agenjo, J., and Blat, J. (2014). 3D graphics on the web: A survey. *Computers & Graphics*, 41, 43–61.
- Gumilang, G. S., and Robinson, M. (2022). Semar: A Personal Model for Counselors. *Pamomong: Journal of Islamic Educational Counseling*, 3(2), 73–84. <https://doi.org/10.18326/pamomong.v3i2.73-84>
- Kim, J. W., Choi, J. Y., Ha, E. J., and Choi, J. H. (2023). Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model. *Applied Sciences (Switzerland)*, 13(4).
- Lin, Y., Jiao, X., and Zhao, L. (2023). Detection of 3D Human Posture Based on Improved Mediapipe. *Journal of Computer and Communications*, 11(02), 102–121.
- 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. (2019). MediaPipe: A Framework for Building Perception Pipelines. *Arxiv*.
- Menolotto, M., Komaris, D.-S., Tedesco, S., O'Flynn, B., and Walsh, M. (2020). Motion Capture

- Technology in Industrial Applications: A Systematic Review. *Sensors*, 20(19), 5687.
- Min, X., Sun, S., Wang, H., Zhang, X., Li, C., and Zhang, X. (2019). Motion capture research: 3D human pose recovery based on RGB video sequences. *Applied Sciences (Switzerland)*, 9(17).
- Odeh, A., and Odeh, N. (2024). OpenCV and its Applications in Artificial Intelligent Systems. *2024 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)*, 242–249.
- Potenziani, M., Callieri, M., Dellepiane, M., and Scopigno, R. (2018). Publishing and consuming 3D content on the web: A survey. *Foundations and Trends in Computer Graphics and Vision*, 10(4), 244–333.
- Salisu, S., Ruhaiyem, N. I. R., Eisa, T. A. E., Nasser, M., Saeed, F., and Younis, H. A. (2023). Motion Capture Technologies for Ergonomics: A Systematic Literature Review. *Diagnostics*, 13(15), 1–16.
- Scott, B., Seyres, M., Philp, F., Chadwick, E. K., and Blana, D. (2022). Healthcare applications of single camera markerless motion capture: a scoping review. *PeerJ*, 10, 1–27.
- Sharma, S., Verma, S., Kumar, M., and Sharma, L. (2019). Use of Motion Capture in 3D Animation: Motion Capture Systems, Challenges, and Recent Trends. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 289–294.
- Wiley, V., and Lucas, T. (2018). Computer Vision and Image Processing: A Paper Review. *International Journal of Artificial Intelligence Research*, 2(1), 22.