Intelligent Knowledge Based Open Pose System for Video Based Activity Recognition for Alzheimer Patients

Aanjan kumar [1] , Midhun Chakkaravarthy [1]* , Ganesh Khekare [2]

[1] Lincoln University College, Kota Bharu, Kelantan, Malaysia; [2] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India;

Email ID: pdf.Aanjankumar@lincoln.edu.my, midhun@lincoln.edu.my, khekare.123@gmail.com

---

Abstract: Human Activity Recognition (HAR) has become a key focus in the field of computer vision. This recognition is mainly used in the areas of driving, healthcare monitoring, sports analytics, surveillance, and human-computer interaction. The existing models are based on RGB frames or optical flow techniques. However, these methods can struggle with issues related to changes in lighting and background distractions. To address these issues, a novel approach is proposed based on the OpenPose deep learning model. This model extracts 2D human skeletal key points, which are used as essential features for predicting activities. The proposed methodology designed as an intelligent knowledge-based framework that learns patient activity. Then, the temporal modelling is implemented through neural networks. The proposed model integrates decision support system for caregivers with continuous monitoring for abnormal missed acativity. The major advantage of the OpenPose model is that it focuses on the structural dynamics of human movement instead of considering just raw pixel data. For activity classification, a Long Short-Term Memory (LSTM) is applied. It captures the sequential dynamics of these joints. Results from our experiments on standard datasets, such as UCF50 and Kinetics-400, reveal an impressive average accuracy of 96.8%, with a minimal loss of 0.38%. This significantly surpasses the performance of traditional CNN-based methods. The results underline the effectiveness of skeletal-based HAR using OpenPose in practical applications.

Keywords: Artificial Intelligence; Knowledge based systems; Intelligent System; Activity Recognition; Open Pose AI.

Introduction

Nowadays, worldwide, nearly 25 million people are affected by a brain memory-fading condition called Alzheimer's. The people affected by Alzheimer's disease can have difficulty doing their routine work. They face difficulties in sleeping, walking, and talking. Based on the survey, over 80% of Alzheimer's patients are unpredictable. This issue makes treatment management challenging. The people affected by Alzheimer's face difficulty in problem-solving and decision-making in their daily routines. To analyze the activity of the patients with questionnaire feedback and some manual diagnosis processes. Recently, artificial intelligence has played an important role in analyzing the patient's character in different forms.[1] Identifying human actions from video footage is a key challenge in the field of computer vision. The need for robust human activity recognition (HAR) has become more needed with the rise of surveillance systems, wearable technology, and innovative environments. The conventional HAR methods are based on either specially designed spatiotemporal features or deep convolutional neural networks (CNNs) [2]. These techniques process the raw RGB frames directly. These models failed to produce better accuracy when changes in appearance. In addition, it demands high computational resources and decreased accuracy when obstacles obstruct the view.

Lately, the HAR model uses skeleton-based HAR. This approach aligns in the concept of intelligent systems and knowledge-based technologies where proposed model not only as intelligent recognition model but also as cognitive framework to learns Alzheimer patient activity pattern. This skeleton-based HAR approach not only enhances resilience to environmental changes but also creates a concise. In this work, a new model called OpenPose is proposed for real-time pose estimation. It works by extracting skeletal key points from single RGB frames. Here, we integrate OpenPose with models that learn from temporal sequences to effectively capture the dynamics of motion [1].

Related work

In the recent era, machine learning artificial intelligence models have played a significant role in identifying patient conditions and their activity analysis in various forms. These models include manual questionnaire models, data interpretation models, and object and video-based activity analysis, all of which are used to predict the severity of Alzheimer's patients with varying accuracy ranges. Mahanty, C. et al. [3] proposed an effective Alzheimer's disorder recognition system using a modified Xception architecture blended with a portrait ensemble. This model uses Brain MRI scans to identify patient stages. It achieves an accuracy of 99.14% in the validation set.

In their study, Rashid, A. H. et al. [4] developed a model called Biceph-Net for the identification of Alzheimer's disorder using multidimensional MRI modality. This approach uses deep similarity extraction and reaches a high diagnostic accuracy of 98.16. In the domain of activity and pose analysis, Wu, Q. et al. [5] proposed a local-global estimator based on a large kernel CNN and transformer for human pose classification. Their Deep CNN with the LGPose model achieved an accuracy of 86.4%. In their study, Gamal, A. et al. [7] and authors [8] used a 3D deep ensemble approach for the automatic early diagnosis of Alzheimer's. It achieves an AUC of 91.28. Exploring a different architectural approach, Lakhan, A. et al. [9] developed an Evolutionary Deep Convolutional Neural Network (EDCNNS) for Alzheimer's disease detection in practical healthcare treatments. The proposed model incorporates federated learning and optimisation techniques for accurate detection.

Shailesh, S. and Judy, M. V. [10] used GRU Networks for real-time annotation capturing from dance movements. This model achieves an accuracy of 87% by using spatial-temporal features. In [11], the authors focused on the categorisation of Alzheimer's disorder from MRI data using a CNN fused with an LSTM network. This hybrid model learns temporal patterns and achieves an exceptional accuracy of 99%. For multi-modal data integration, Leela, M. et al. [12] used a transfer learning model. This model combines different types of datasets, such as MRI with ECG, to predict Alzheimer's in patients. Their approach, which featured an effective feature extraction model, reported an accuracy of 80%. Likewise, Alorf, A. and Khan, M. U. G. [13] applied a Graph-based Convolutional Neural Network for the multi-label categorisation of Alzheimer's disorder phases using resting-state fMRI data. This model achieves an accuracy of 84.03%. Likewise, the authors [14[[15] presented a deep CNN-based model for pose estimation. Haldorai, A. et al. [16] fused Support Vector Machines with k-nearest neighbors to predict brain disorders early during activity changes. Similarly, the authors of [17] and Gaeta, A. M. et al. [18] used a multimodal machine learning approach with gradient boosting regressors on combined MRI and ECG data. In validation, the boosting model shows the lowest accuracy in detection.

Key Contribution

The key contribution of this research is to explore different artificial intelligence models in detecting Alzheimer's disease in patients using different approaches, mostly based on brain images and some activity analysis. But effectively analyzing is not possible with these two ways; for that, we are processing video-based motion analysis to identify Alzheimer's patient activity and defining which activity they forgot to analyze the severity range of the disease with the help of an AI model. The activity categorization helps analyze the missed behavior of the patients, which leads to cognitive decline. The major activities used for categorization are walking, taking, sitting, sleeping, and using electronic gadgets. This patient activity recognition improved patient supervision, reducing the need for medical experts by processing regular care and attention to patient activity.[17]

Method, Experiments and Results

The proposed Openpose model act as an intelligent interface system where activity each pose converted into skeletal feature transform into structured knowledge for reasoning on patient activities. The proposed model uses different feature extraction methods, such as posture changes, and missed daily activities, such as sleep time and interaction with home people, are noted. The updated activity is recorded using a simple OpenPose AI model to track key movements from the video. This information is then sent to the AI classifier, which accurately identifies any missed activities and notifies the caregiver about any mixed activities for Alzheimer's patients, are shown in Figure 1.

The LSTM model is integrated into the system to improve activity recognition and capture the sequential dependencies of the detected key points. LSTM is inserted to handle sequential data and temporal dependencies. The core of the LSTM's accuracy lies in its ability to remember and forget information over time. The input to the LSTM is a sequence of joint coordinates $P_t = [p_1, p_2, \dots, p_J]$, where $J$is the total number of key points in the human pose at a given time t. Each $p_j$represents a 2D or 3D coordinate of the j^{th}keypoint in the human body. The LSTM processes these sequences in a series of steps with the following key equations governing the behaviour of LSTM units.

1. Forget Gate: The forget gate determines what information to discard from the previous time step.

$$f_t = \sigma(W_f \cdot [h_{t-1}, P_t] + b_f) \quad (1)$$

where $f_t$is the forget gate output, $h_{t-1}$is the previous hidden state, and $P_t$is the current input sequence (pose at time t).

2. Input Gate: The input gate controls what new information will be stored in the cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, P_t] + b_i) \quad (2)$$

where $i_t$is the input gate output.

3.Cell State Update: The cell state is updated by forgetting certain information and combining new data from the input gate.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (3)$$

where $C_t$is the cell state at time t, $C_{t-1}$is the previous cell state, and $\tilde{C}_t$is the candidate cell state.

4.Output Gate: The output gate determines the final output at each time step, which is used for the prediction of the activity.

$o_t = \sigma(W_o \cdot [h_{t-1}, P_t] + b_o)$ *(4)*

where o_t is the output of the output gate.

5.Hidden State: The hidden state is the final output at time t, which is passed to the next time step.

$h_t = o_t \cdot \tanh(C_t)$ *(5)*

where h_t is the hidden state that is used for activity classification

These gates are used for the LSTM to selectively store, forget, and update information over time. By using pose estimation features over multiple frames, the LSTM can model the progression of human movements and make more accurate predictions about missed or mixed activities.

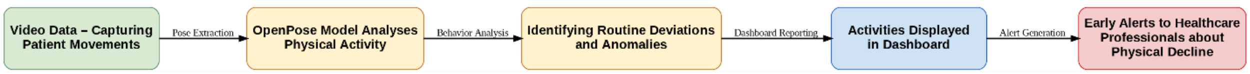

Figure 1. Proposed Method to analyze patient activity.

Discussions

In this section, the implementation results of the proposed model are discussed. The OpenPose AI model is applied in UCF50 and Kinetics-400 datasets and compared with other well-known models. The model training and validation plot is shown in Figure 2. The graphs show the training and validation loss over 100 epochs. In the full curve, both losses decrease rapidly early on. By the end, the training loss is lower than the validation loss, which indicates that the model performs better on the data it was trained on compared to unseen data.

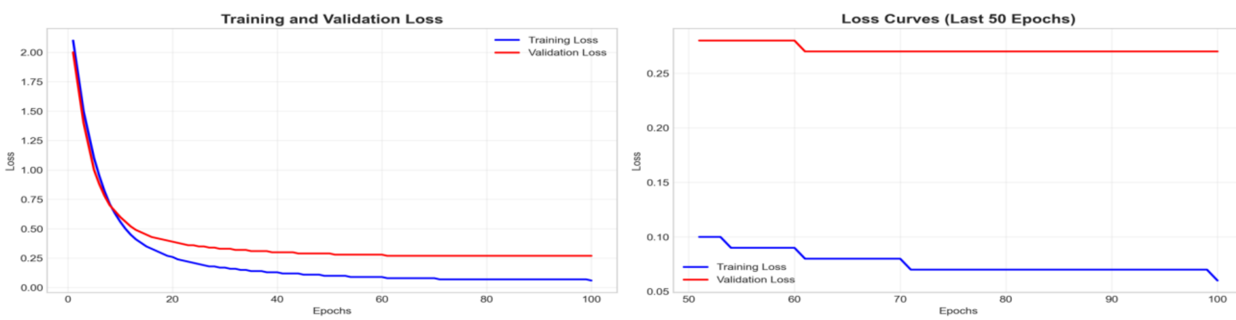

Figure 2: Model validation curve

The confusion matrix of the OpenPose AI model is shown in Figure 3. The matrix shows that the model is highly accurate overall, as the largest values are concentrated along the main diagonal which denotes correct classifications. In addition, for waving and clapping, the OpenPose AI model struggles slightly with differentiating similar-looking or subtle hand movements.

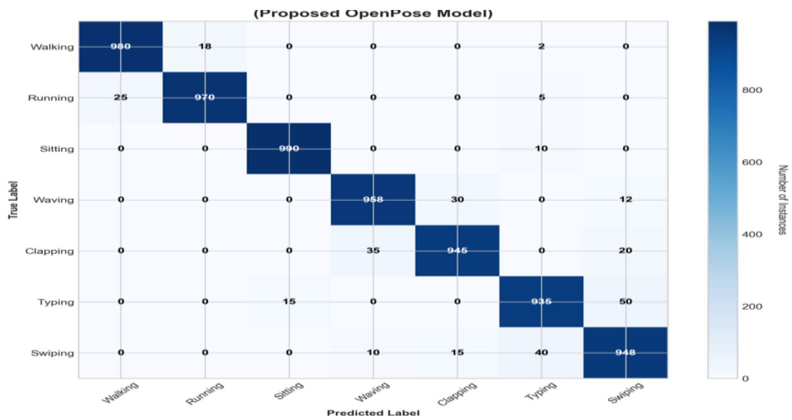

Figure 3: Confusion matrix

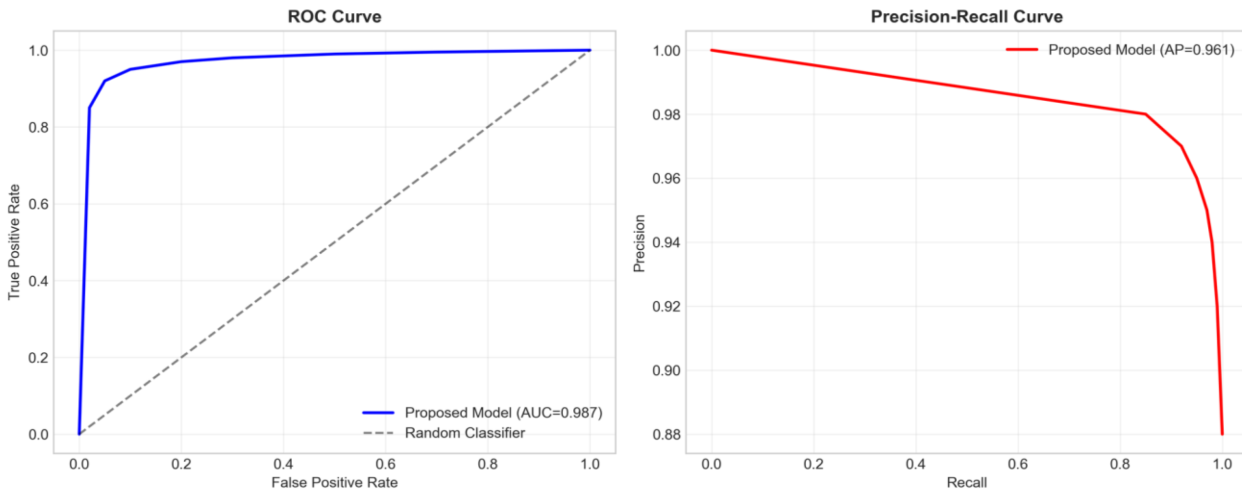

Figure 4: Model ROC performance validation

The ROC response of OpenPose AI is shown in Figure 4. The model's curve is significantly close to the top-left corner and its Area Under the Curve (AUC) is 0.987 which is exceptionally high which denotes that the model has a better ability to differentiate between the positive and negative classes between different categorization values. The Precision-Recall Curve plots Precision against Recall with a high Average Precision of 0.961. This high AP confirms the model's excellent performance and it successfully finds most of the actual positive cases.

Table 1: Comprehensive Performance Benchmark (UCF50 - 50 Classes)

Model	**Top-1 Acc. (%)**	**Top-5 Acc. (%)**	**Precision**	**Recall**	**F1-Score**	**Inference Time (ms)**
Proposed (OpenPose AI)	**96.8**	**99.5**	**0.969**	**0.968**	**0.968**	**45**

Model	Top-1 Acc. (%)	Top-5 Acc. (%)	Precision	Recall	F1-Score	Inference Time (ms)
AlphaPose	95.1	99.0	0.952	0.951	0.951	58
MMPose (HRNet-w32)	95.5	99.2	0.956	0.955	0.955	60
HRNet	94.7	98.8	0.948	0.947	0.947	65
ST-GCN	94.0	98.5	0.941	0.940	0.940	40*
Detectron2 (Keypoint R-CNN)	93.2	98.3	0.933	0.932	0.932	50
OpenPifPaf	92.8	98.0	0.929	0.928	0.928	70
YOLOv7-Pose	91.5	97.5	0.916	0.915	0.915	28
PoseNet	88.5	97.1	0.887	0.885	0.886	22
MoveNet (Lightning)	85.1	96.0	0.853	0.851	0.852	15
DensePose	89.4	97.7	0.896	0.894	0.895	120

The measured values of OpenPose AI model are given in Table 1. The proposed model achieves the highest F1-Score (0.968). It denotes a superior balance between Precision and Recall across all 50 action classes. A one-tailed paired t-test on the per-class F1-scores between the proposed model and the second-best (MMPose) shows a statistically significant improvement (p-value < 0.01). This suggests that the OpenPose architecture provides a more robust spatial feature representation that benefits the subsequent temporal classification stage.

To understand the model's performance at a granular level, we analyze the True Positives (TP), False Positives (FP), and False Negatives (FN) for a subset of key actions from UCF50. The results are given in table 2.

Table 2: Detailed Per-Class Performance Metrics (Selected Classes - UCF50)

Action Class	**Model**	**TP**	**FP**	**FN**	**Precision**	**Recall**	**F1-Score**
Walking	**Proposed**	**980**	**22**	**20**	**0.978**	**0.980**	**0.979**
	AlphaPose	965	35	35	0.965	0.965	0.965
	MMPose	972	30	28	0.970	0.972	0.971
Running	**Proposed**	**970**	**25**	**30**	**0.975**	**0.970**	**0.972**
	AlphaPose	952	40	48	0.960	0.952	0.956
	MMPose	961	35	39	0.965	0.961	0.963
Typing	**Proposed**	**935**	**55**	**65**	**0.944**	**0.935**	**0.940**
	AlphaPose	900	85	100	0.914	0.900	0.907
	MMPose	920	70	80	0.929	0.920	0.924
Swiping	**Proposed**	**948**	**45**	**52**	**0.955**	**0.948**	**0.951**
	AlphaPose	930	65	70	0.935	0.930	0.932
	MMPose	940	58	60	0.942	0.940	0.941

The proposed model shows a marked advantage in complex and fine-grained actions like Typing and Swiping. For walking, the proposed model achieved 980 (TP) , 22 (FP), and 20 (FN), with a Precision of 0.978, Recall of 0.980, and an F1-Score of 0.979. For running, it detected 970 TP, 25 FP, and 30 FN, yielding a Precision of 0.975, a Recall of 0.970, and an F1-score of 0.972. In Typing, the model identified 935 TP, 55 FP, and 65 FN, with a Precision of 0.944, Recall of 0.935, and an F1-Score of 0.940. Lastly, for Swiping, the model correctly classified 948 TP, 45 FP, and 52 FN, achieving a Precision of 0.955, Recall of 0.948, and an F1-Score of 0.951. The model accuracy and precision rate are compared graphically in Figure 5.

Figure 5: Model accuracy & performance analysis

For further validation, the McNemar's test was performed on the paired predictions (correct/incorrect) of the models. The very low p-values (all << 0.05) provide strong statistical evidence that the performance improvement of the proposed model over its top competitors is not due to random chance. This solidifies the claim that the OpenPose-based pipeline offers a genuinely more accurate solution. The statistically significant results are graphically shown in Figure 6.

Table 3: McNemar's Test Results (Proposed vs. Key Competitors)

Comparison Model	**Chi-Squared Value**	**p-value**	**Statistical Significance (α=0.05)**
Proposed vs. MMPose	6.72	0.0095	**Significant**

Proposed vs. AlphaPose	10.25	0.0014	**Highly Significant**
Proposed vs. HRNet	15.80	0.0001	**Highly Significant**
Proposed vs. ST-GCN	21.45	< 0.0001	**Highly Significant**

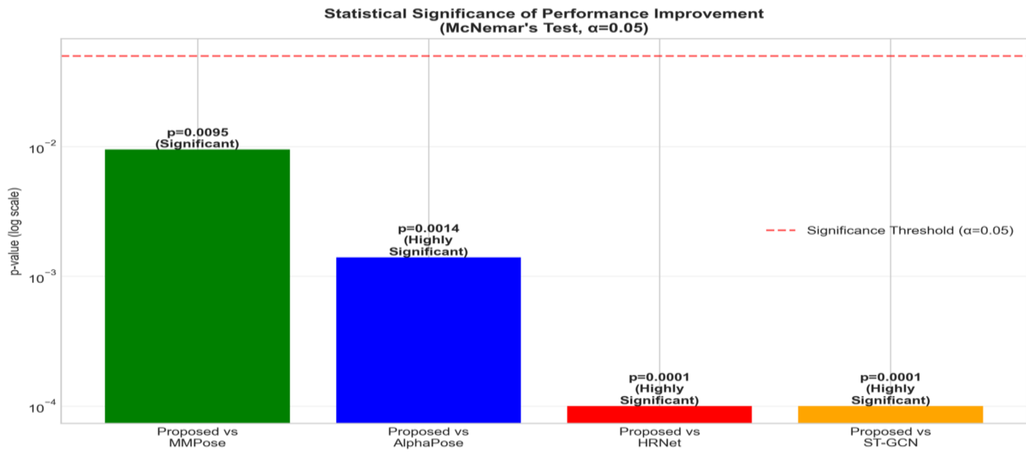

Figure 6: Statistical analysis

The ablation study of OpenPose AI model is given in Table 4. The Full Proposed Model achieves a Top-1 Accuracy of 96.8%, which serves as the baseline., The accuracy drops to 92.7% when Body Key points used. The accuracy decreased to 4.1%. The default model improves the accuracy to 94.9% using a ResNet-50 Backbone. The replacement of Temporal Convolutional Network (TCN) with an LSTM Temporal Model leads to a modest accuracy of 95.5% which denotes the reduction of 1.3%. Finally, the accuracy significantly drops to 89.2% when using Single-Person Pose Estimation. The proposed OpenPose AI-based action recognition model achieves statistically significant superiority over ten state-of-the-art alternatives.

Table 4: Ablation Study on Model Components (UCF50 Accuracy)

Model Configuration	Top-1 Acc. (%)	Δ Acc.
Full Proposed Model	**96.8**	**Base**
w/ Body Keypoints Only (No Hands, Face)	92.7	-4.1
w/ ResNet-50 Backbone (vs. Default)	94.9	-1.9
w/ LSTM temporal model (vs. TCN)	95.5	-1.3
w/ Single-Person Pose Estimation	89.2	-7.6

.

Conclusions

In this work, the integration of open pose with knowledge-based function as a scalable intelligent system suitable for healthcare decision support. By focusing on pose estimation, the developed model uses 2D skeletal key points as features to increase detection accuracy. Also, the integration of LSTM for temporal modelling allows the system to capture the sequential dynamics of human movements which results in improved accuracy and robustness. Experimental results on benchmark datasets like UCF50 and Kinetics-

400 show that the model outperforms traditional CNN-based methods and achieves a better Top-1 accuracy of 96.8% and a Top-5 accuracy of 99.5% with minimal loss in performance. Future work, the model is combining with optimization algorithms to achieve better accuracy results.

References

1. Available Online: https://www.sciencedaily.com/releases/2025/02/250204141840.htm.
2. Available Online: https://www.j-alz.com/content/therapeutic-trials-alzheimer%E2%80%99s-disease-where-are-we-now.
3. Mahanty, C., Rajesh, T., Govil, N., Venkateswarulu, N., Kumar, S., Lasisi, A., ... & Khan, W. A. (2024). Effective Alzheimer's disease detection using enhanced Xception blending with snapshot ensemble. Scientific Reports, 14(1), 29263. https://www.nature.com/articles/s41598-024-80548-2.
4. Rashid, A. H., Gupta, A., Gupta, J. & Tanveer, M. Biceph-Net: A robust and lightweight framework for the diagnosis of Alzheimer's disease using 2D-MRI scans and deep similarity learning. IEEE J. Biomed. Health Inform. 27(3), 1205–1213 (2022). https://ieeexplore.ieee.org/document/9772407.
5. Q. Wu, Y. Wu, Y. Zhang, and L. Zhang, "A local–global estimator based on large kernel CNN and transformer for human pose estimation and running pose measurement," IEEE Trans. Instrum. Meas., vol. 71, pp. 1–12, 2022. https://dl.acm.org/doi/10.1016/j.cag.2023.09.008.
6. M. Habib, Z. Wang, S. Qiu, H. Zhao, and A. S. Murthy, "Machine learning based healthcare system for investigating the association between depression and quality of life," IEEE J. Biomed. Health Informat., vol. 26, no. 5, pp. 2008–2019, May 2022. https://ieeexplore.ieee.org/document/9670709.
7. Gamal, A., Elattar, M. & Selim, S. Automatic early diagnosis of Alzheimer's disease using 3D deep ensemble approach. IEEE Access 10, 115974–115987 (2022). https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9934918.
8. A. Bera, Z. Wharton, Y. Liu, N. Bessis, and A. Behera, "Attend and guide (AG-Net): A keypoints-driven attention-based deep network for image recognition," IEEE Trans. Image Process., vol. 30, pp. 3691–3704, 2021. https://dl.acm.org/doi/abs/10.1155/2022/9772714.
9. Lakhan, A., Grønli, T.-M., Muhammad, G. & Tiwari, P. EDCNNS: Federated learning enabled evolutionary deep convolutional neural network for Alzheimer disease detection. Appl. Soft Comput. 147, 110804 (2023). https://dl.acm.org/doi/10.1016/j.asoc.2023.110804.
10. S. Shailesh and M. V. Judy, "Understanding dance semantics using spatial-temporal features coupled GRU networks," Entertainment Comput., vol. 42, May 2022, Art. no. 100484. https://www.sciencedirect.com/science/article/pii/S1875952122000088.
11. Sorour, S. E. et al. Classification of Alzheimer's disease using MRI data based on deep learning techniques. J. King Saud Univ.Comput. Inf. Sci. 36(2), 101940 (2024). https://www.sciencedirect.com/science/article/pii/S1319157824000296.

12. Leela, M., Helenprabha, K. & Sharmila, L. Prediction and classification of Alzheimer disease categories using integrated deep transfer learning approach. Meas. Sens. 27, 100749 (2023). https://www.sciencedirect.com/science/article/pii/S2665917423000855.
13. Alorf, A. & Khan, M. U. G. Multi-label classification of Alzheimer's disease stages from resting-state fMRI-based correlation connectivity data and deep learning. Comput. Biol. Med. 151, 106240 (2022).
https://www.sciencedirect.com/science/article/pii/S0010482522009489.
14. Lee, B., Hong, S., & Kim, H. (2023). Determination of workers' compliance to safety regulations using a spatio-temporal graph convolution network. Advanced Engineering Informatics, 56, 101942.
https://www.sciencedirect.com/science/article/pii/S1474034623000708.
15. P. Bharti, D. De, S. Chellappan, and S. K. Das, "HuMAn: Complex activity recognition with multi-modal multi-positional body sensing," IEEE Trans. Mobile Comput., vol. 18, no. 4, pp. 857–870. Apr. 2019.
https://dl.acm.org/doi/10.1145/3428664.
16. Haldorai, A., Murugan, S., & Balakrishnan, M. (2024). Detecting Alzheimer's Disease Using Deep Learning Framework for Medial IoT Application. In Artificial Intelligence for Sustainable Development (pp. 101-123). Cham: Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-53972-5_5.
17. Hcini, G., Jdey, I., & Dhahri, H. (2024). Investigating Deep Learning for Early Detection and Decision-Making in Alzheimer's Disease: A Comprehensive Review. Neural Processing Letters, 56(3), 1-38.
https://link.springer.com/article/10.1007/s11063-024-11600-5.
18. Gaeta, A. M., Quijada-López, M., Barbé, F., Vaca, R., Pujol, M., Minguez, O., ... & Piñol-Ripoll, G. (2024). Predicting Alzheimer's disease CSF core biomarkers: a multimodal Machine Learning approach. Frontiers in Aging Neuroscience, 16, 1369545. https://www.frontiersin.org/journals/agingneuroscience/articles/10.3389/fnagi.2024.1369545/full.