Advancing Healthcare through XAI and Motion Analysis Integration

Thanh Nguyen

December 13, 2023

Abstract

In recent years, artificial intelligence (AI) has become increasingly important in various domains in our modern society. However, its lack of complete accuracy presents challenges, especially in critical areas like healthcare where precision is crucial. Motion analysis has emerged as a promising solution that can enhance the accuracy of AI applications in healthcare. By incorporating advanced motion tracking technologies, artificial intelligence (AI) systems have the ability to extract intricate information from movements, thereby enhancing the dependability of healthcare AI in diagnostic procedures.

1 Introduction

In an era filled with some of the most advancing technology, the integration of cutting-edge technologies remains both a promise and a challenge. Explainable Artificial Intelligence (XAI), while holding tremendous potential for enhanced decision-making, often grapples with the intricacies of transparency and interpretability. Amid this uncertainty, the field of motion analysis emerges as a pivotal ally, offering a tangible pathway to augment XAI's decision-making capabilities.By analyzing intricate patterns in human motion, motion analysis acts as a supplementary instrument that can aid in the improvement of XAI algorithms. This collaboration presents fresh opportunities for healthcare applications, leading to advancements in diagnostic techniques, tailored treatment plans, and visualizes a future where ambiguity is replaced by accuracy in healthcare decision-making.

2 Method

This analysis takes a dive into the powerful combination of Explainable Artificial Intelligence (XAI) and motion analysis, focusing on the OpenPose algorithm. While we delve into the emerging possibilities of XAI, specifically in terms of transparency and trust in artificial intelligence systems, we also analyze the strong capabilities of OpenPose in extracting detailed human motion data.

2.1 XAI Potential in Healthcare

Mental health challenges are highlighted, noting the difficulty in direct observation and the impact on overall well-being, with depression cited as a leading cause of lost working hours globally which leads to conditions affecting the way people think, feel and behave...poor quality of life, degenerated physical health [1]. The text emphasizes the complexity of mental disorders, their early onset, and the challenges in patient responsiveness such as heterogeneous, dynamic, and multi-causal phenomena and patients tend to be less responsive and compliant [1]. It introduces the potential of Artificial Intelligence (AI) and Machine Learning (ML) to provide clinical decision support, particularly in analyzing diverse datasets for early detection and diagnosis of disorders. Individualized medical data has enabled technologies such as Artificial Intelligence (AI) and Machine Learning (ML) which is employed to analyze big and diverse data to identify patterns that associate mental disorders with clinical data, biometrics, behaviors, and social interactions" [1].

Over the past two decades, there has been a specific emphasis on utilizing technology, specifically machine learning (ML), to assist in the identification and evaluation of mental health concerns. The direct observation of indicators related to mental health can often be a complex task. However, by applying ML techniques to brain scans and medical data, it becomes possible to obtain "proxy measures" that can provide valuable insights into brain-related health issues [2]. We also acknowledges the limitations of black box ML models, which lack interpretability, hindering understanding and trust in mental health applications. As expected, complex models tend to outperform simpler ones, however, these highperformant models come with the trade-off of black box functionality [2]. Explainable Artificial Intelligence (XAI) can be used as a solution to address these drawbacks, aiming to enhance transparency, trust, and understanding among medical experts while improving mental health interventions. Explainable Artificial Intelligence (XAI) approaches aim to address these drawbacks. Without explainability, developed methods are incapable of devising new theories and leading to incremental science. For instance, a recent review pointing to pitfalls and misconducts in the proposals of new DL approaches may represent this state of affairs [3].

Explainable Artificial Intelligence (XAI) is a research focus aiming to enhance transparency in AI systems, particularly deep learning models. It ensures accountability, transparency, and model improvement. XAI benefits various applications like medical diagnosis, employing techniques such as visualization for mammography image analysis [2]. The text emphasizes the importance of explainability in deep learning, preventing pitfalls and misconduct. It explores the rule extraction technique, transferring feature maps from a CNN to a Discretized Interpretable Multi-Layer Perceptron for enhanced interpretability. Challenges in XAI are acknowledged, prompting the exploration of hybrid approaches and collaboration with clinicians for real-world evaluations. The ultimate goal is to integrate AI technologies into clinical pathways for improved diagnosis and treatment processes [2].

However, there are several challenges that need to be addressed. One of the challenges is the influence of intricate correlations among diagnostic features, which can make it difficult to interpret the results accurately. Additionally, the current interpretability methodologies have their limitations, further complicating the process [2]. To overcome these challenges, researchers are exploring hybrid approaches, particularly neurosymbolic ones. These approaches combine the strengths of neural networks and symbolic reasoning to enhance the interpretability of AI models in mental health diagnosis. From an understanding of human brain development [4] to analyzing biomarkers for AD [5, most would agree that this approach is very useful for adding to the knowledge accumulated so far. Nevertheless, while the solutions this brings are promising, at present we are still a long way from achieving them [8].

2.2 Motion Analysis

Motion analysis combines biomechanics, computer vision, and AI to analyze human or object movement, using technologies like motion capture systems and algorithms such as Openpose. It is crucial in sports science, rehabilitation, and ergonomics, enhancing performance, preventing injuries, and conducting biomechanical assessments. AI integration automates the extraction of movement parameters, revolutionizing our understanding of human motion.

2.2.1 Openpose Technology

The research employs the Openpose algorithm to facilitate the identification of anatomical keypoints from video footage, with the primary objective of converting this keypoint data into clinically significant health indicators. Openpose is utilized to process video inputs, generating output data on 25 body keypoints and 21 finger/hand keypoints. The experimental focus centers around six fundamental movements: bow squat, full squat, lying down and raising the feet, sitting and turning at the waist, touching the back with hands from top to bottom, and touching the back with hands from bottom to top. Ten health indicators, classified as primary and secondary, are established to clinically evaluate the movements. These indicators encompass angles and distances relevant to torso stability, joint mobility, and movement accuracy. The process entails meticulous configuration and calculation of health indicators, taking into account primary and secondary observations. Following the application of the Openpose algorithm, the study progresses through film pre-processing, Openpose image processing, keypoint processing, and a final comparison of results with manual measurements, thereby demonstrating the feasibility of Openpose for detecting and analyzing physical health.

- **Film Pre-processing**

  An initial stage where video inputs undergo transformations, including flipping footage for specific movements like lying down, raising the feet, and touching the back. This step is critical for standardization, ensuring optimal detection accuracy in subsequent stages.

  Sample Collections

  Bow squat, Full squat, Lying down and raising the feet, Sitting and turning at the waist, Touching the back with hands from top to bottom, Touching the back with hands from bottom to top.

- **Openpose Image Processing**

  Openpose identifies 25 body keypoints and 21 finger/hand keypoints from the processed videos. These keypoints serve as foundational data for the subsequent calculation of health indicators, forming the basis for further analysis of human movements [6].

Body Keypoints

In total, there are 25 body keypoints (as in *Figure 1*), including ears, eyes, nose, the center point of the shoulders and limb joints (shoulder, elbow, wrist, hip, knee and ankle). The data used in this research mainly relate to body keypoints, and these were used for the purposes of detecting the location of limbs.

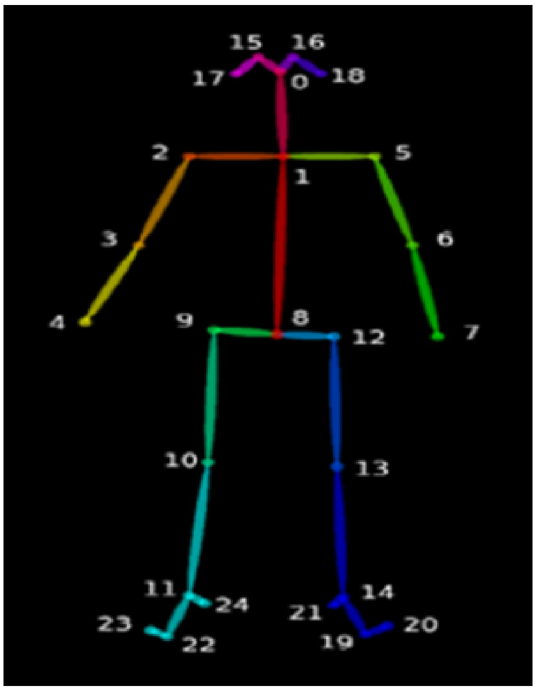

Figure 1: The 25 keypoints recognized by Openpose.

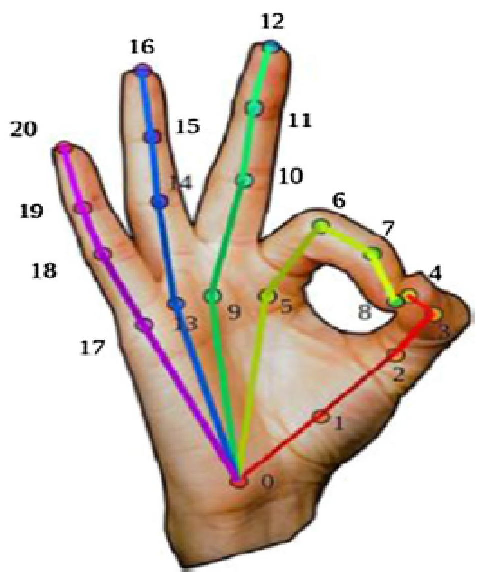

Figure 2: The 21 keypoints recognized by Openpose.

Finger/Hand Keypoints
In total, there are 21 keypoints in the hand (as in *Figure 2*), including the three joints of each finger, the fingertips and the heel of the palm. This paper is intended to be used as a reference point for specific movements and limb keypoints. The point mix is used as the reference basis for completion of the action [6].

- **Keypoint Processing**
  Involving the establishment of ten health indicators, categorized as main and secondary, for six fundamental movements. These indicators assess various aspects, including torso stability, joint mobility, and movement correctness. The keypoints identified by Openpose play a crucial role in quantifying these indicators, facilitating a nuanced evaluation of an individual's physical health.
  Setting and calculation of health indicators
  The process for establishing health indicators begins with using the Openpose algorithm on video recordings. This yields data on 25 body keypoints and 21 finger/hand keypoints. These keypoints are used to derive clinical health indicators, following clinical judgment standards. The subsequent calculations are crucial for meaningful clinical interpretation [6].
  Step 1 (Bow Squat): This involves assessing the stability of the torso during a bow squat. Key indicators include the angle between specific lines on the left and right sides, focusing on the vertical axis and foot clamps.
  Step 2 (Full Squat): The primary indicator for a full squat is the minimum knee angle, which provides important information about foot stiffness and joint mobility. When assessing the squat, other factors such as knee stability and the angle between the calf bone and forearms are also considered.
  Step 3 (Lying down and raising the feet: foot angle): The foot angle, determined by specific lines, is the primary measure in this stage. It reflects joint flexibility and mobility. Knee flexion angles also contribute to movement quality assessment.
  Step 4 (Sitting and turning at the waist: rotation angle): This step focuses on detecting the rotation angle of the shoulder during sitting and turning at the waist. It assesses the stiffness of the back and hips, as well as the stability of the hip joint, crucial for accurate testing posture.
  Step 5 (Touching the back with hands from top to bottom: back-touching position): Standards are set using shoulder keypoints, considering variations in camera distance and height. Stiffness and shoulder injuries are considered in result interpretation.

Step 6 (Touching the back with hands from bottom to top: back-touching position): Similar to the previous step, this assesses the back-touching position but from bottom to top. Standards are applied based on shoulder keypoints, considering variations in camera distance and height, while accounting for stiffness and shoulder injuries.

Rotating the left and right hand (foot)
The accuracy of data obtained from subjects lying flat and performing movements was affected by the rotation of hand and foot keypoints. An unexpected data bounce during video detection led to further investigation, revealing errors on both sides despite successful detection of limb keypoints. This challenge was more prominent when the subject's face was not captured and they wore black pants, due to issues with face direction and light-shadow dynamics. To overcome this, the left and right keypoint coordinates were adjusted to ensure accurate alignment with the designated action standard. This adjustment was necessary for movements like lying down, raising the feet, and touching the back, as misalignment of keypoints could lead to inaccuracies in the final health index. Detection and correction of misaligned keypoints were crucial to accurately represent the intended movements and mitigate their impact on the overall assessment [7].

Removing undetected points and offset points
Keypoint detection faced challenges, especially when subjects went from lying flat to a 90-degree flip, causing occasional offsets. This problem was particularly noticeable in detecting feet movements, leading to detection errors. The concern was the potential for sudden signal jumps, which could greatly affect the accuracy of health indicators. By removing undetected and offset points, the values became more consistent with the actual situation, facilitating a clearer understanding of the maximum values. This comprehensive keypoint processing procedure was indispensable in refining the data, ensuring that health indicators accurately reflected the subjects' movements and minimizing errors attributable to detection challenges [6].

- **Comparison of Results**
  The health indicators calculated by the system are rigorously evaluated against manual measurements conducted by physical therapists. This step serves as a validation process, assessing the accuracy and reliability of the Openpose-based system in detecting and analyzing physical health, providing insights into its clinical applicability.

2.2.2 Openpose Fall Detection

The extraction of key points using OpenPose is prone to bias, which can adversely affect the accuracy of fall detection. To address this issue, the proposed algorithm combines OpenPose and MobileNetV2 to improve the extracted features and minimize detection errors resulting from biased key point extraction.

- **Features enhancement**

  The section on enhancing features addresses the challenges faced when using OpenPose for extracting key points in fall detection. The presence of varying light conditions in the video data can introduce bias in the extraction of key points, thereby affecting the accuracy of detection. To tackle this issue, OpenPose performs key point extraction, annotates the original image, and utilizes these annotated images as input for fall detection with MobileNetV2. This process makes use of pose features from the original image, which serve as a reference for subsequent classification detection. The method involves preprocessing images, employing VGG-19 for feature extraction, predicting position confidence maps and part affinity fields, optimizing network parameters, and marking key points in the original image [7]. The process primarily focuses on connecting body parts, starting from the torso, and utilizes colorful annotations to differentiate human torsos. This approach aims to enhance feature extraction and enhance the accuracy of fall detection by incorporating pose features from the original image.

- **Fall Detection**

  The primary focus of the fall detection section is to enhance the accuracy of fall detection by utilizing the lightweight neural network MobileNetV2. In this study, MobileNetV2 is fine-tuned and employed for fall classification detection. A comparative experiment is conducted with Efficientnet and EfficientnetV2, revealing that MobileNetV2 achieves higher accuracy in fall detection on both the Le2i dataset (98.5% compared to 93.5% and 94.92%) and the UR dataset (96.3% compared to 95.93% and 96%) [7].

  To address specific challenges, the network framework is modified by incorporating additional components. The original MobileNetV2 framework is extended with a fully connected layer and a softmax classifier to ensure compatibility with the requirements of fall detection. Furthermore, to enhance the network's learning ability, a Convolutional Block Attention Module (CBAM) attention mechanism is integrated at the beginning of the network [7]. The overall architecture of the network includes convolutional layers, the CBAM attention framework, and seven bottlenecks with different blocks. To slow down the drop in feature dimension, a fully connected layer is introduced at the end of MobileNetV2, followed by the softmax function for probability calculation [7].

For parameter optimization, the fall detection model employs the cross-entropy loss function and the Adam algorithm (*Figure 3*), where m is the number of samples, y_t is the real category, y_p is the prediction category, q is the sample.

$$Loss = -\frac{1}{m}\sum_{q}\left[y\_t \cdot \ln(y\_p) + (1 - y\_t)\ln(1 - y\_p)\right],$$

Figure 3: Cross-Entropy Loss function..

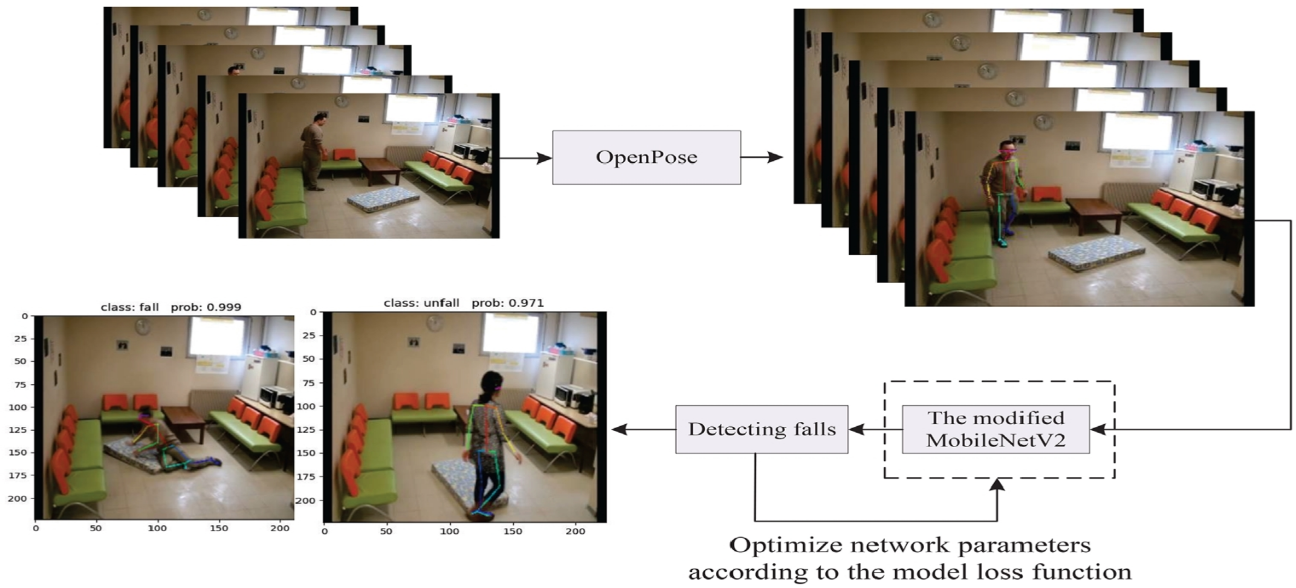

Figure 4: Overall Flow Framework.

Figure 5: Brightening Process.

In order to enhance the accuracy of keypoint extraction, we use equation *Figure 6* to adjust the brightness of excessively dark images. By determining the average brightness value of typical images from different scenes, we establish a threshold of 130.7 for image brightness. By calculating the pixel value distribution, we found that the average value of most image pixels is between 39 and 56. In order to contain more pixel information, we set the pixel existence interval to [1,99], discarding any values outside this range [7]. Subsequently, we normalize the image pixels by mapping them back to the [0,255] interval, ensuring that outliers are removed. To prevent pixel overflow, we set the interval to [255*0.1,255*0.9]. Here, x_o represents the original pixel value, while x_p represents the highlighted pixel value.

$$\frac{x\_o}{99-1} = \frac{x\_p}{\left[255 \times (0.9-0.1)\right]}.$$

Figure 6: Brightening Formula.

The final output, which indicates whether a fall has occurred or not, along with its corresponding probability, is determined through deep separable convolution, average pooling, and fully connected layers, ultimatelyyields great accuracy with 98.60% and 99.75% for Le2i and UR datasets respectively.

- **Overall Framework**
  Step 1: inputs the preprocessed image into the OpenPose network.
  Step 2: extracts features.
  Step 3: predicts the position confidence map and part affinity fields domain of the limbs, while optimizing the network model parameters based on the loss function.
  Step 4: associates body parts with the human body and marks the image with key points. These four steps utilize the Openpose algorithm to label human key points in the picture.
  Step 5: inputs images with key point information into MobileNetV2 and extracts features using the deep convolution module.
  Step 6: employs the Adam algorithm to optimize the model parameters according to the loss function, enabling fall detection.

3 Conclusion

Explainable Artificial Intelligence (XAI) has the potential to greatly enhance transparency and trust in AI systems, especially in complex models like deep learning. In the realm of motion analysis, OpenPose plays a crucial role by extracting human keypoints from images in a way that can be easily interpreted. In the proposed fall detection algorithm, OpenPose is combined with a modified MobileNetV2 network to detect falls based on keypoint and pose information. This integration not only improves the accuracy of features without increasing the complexity of the images, but also corrects any labeling errors that may occur. By utilizing OpenPose in motion analysis, XAI enables a clearer understanding of the decision-making processes in intricate models, providing benefits such as increased confidence, error analysis capabilities, and the potential for model refinement. The collaboration between OpenPose and XAI, as exemplified in fall detection, demonstrates how interpretable insights from motion analysis can contribute to more reliable and transparent AI systems, opening up possibilities for applications in various domains, including healthcare.

References

- Kirkpatrick, Keith. ”Artificial Intelligence and Mental Health: How AI Can Be Used to Improve Diagnosis of Mental Health Conditions.” *Communications of the ACM*, vol. 65, no. 5, May 2022, pp. 32–34. *EBSCOhost*, https://doi.org/10.1145/3524017.

- Antoniou, Grigoris, et al. “Mental Health Diagnosis: A Case for Explainable Artificial Intelligence.” *International Journal on Artificial Intelligence Tools*, vol. 31, no. 3, May 2022, pp. 1–7. *EBSCOhost*, https://doi-org.libproxy.temple.edu/10.1142/S0218213022410032.

- Górriz, J. M., et al. “Computational Approaches to Explainable Artificial Intelligence: Advances in Theory, Applications and Trends.” *Information Fusion*, vol. 100, December 2023, p. N.PAG. *EBSCOhost*, https://doi-org.libproxy.temple.edu/10.1016/j.inffus.2023.101945.

- Kiani, Mehrin, et al. ”Towards understanding human functional brain development with explainable artificial intelligence: Challenges and perspectives.” *IEEE Computational Intelligence Magazine*, 17.1 (2022): 16-33. *Google Scholar*, https://ieeexplore.ieee.org/document/9679739.

- Lombardi, Angela, et al. ”A robust framework to investigate the reliability and stability of explainable artificial intelligence markers of Mild Cognitive Impairment and Alzheimer’s Disease.” *Brain informatics*, 9.1 (2022): 1-17. *Google Scholar*, https://braininformatics.springeropen.com/articles/10.1186/s40708-022-00165-5.

- Huang, Lingfeng, and Genggeng Liu. “Functional Motion Detection Based on Artificial Intelligence.” *Journal of Supercomputing*, vol. 78, no. 3, Feb. 2022, pp. 4290–329. *EBSCOhost*, https://doi-org.libproxy.temple.edu/10.1007/s11227-021-04037-3.

- Gao, Mengqi, et al. “Fall Detection Based on OpenPose and MobileNetV2 Network.” *IET Image Processing (Wiley-Blackwell)*, vol. 17, no. 3, Feb. 2023, pp. 722–32. *EBSCOhost*, https://doi-org.libproxy.temple.edu/10.1049/ipr2.12667.

- Ghassemi, Marzyeh, Luke Oakden-Rayner, and Andrew L. Beam. ”The false hope of current approaches to explainable artificial intelligence in health care.” *The Lancet Digital Health*, 3.11 (2021): e745-e750. *Google Scholar*, https://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00208-9/fulltext.