



A New Method of Non-Invasive Diagnosis for Human Body Structure Using Pose-Detection

I Ketut Swardika¹, Putri Alit Widyastuti Santiary²,
Dewa Ayu Indah Cahya Dewi³, and I Gusti Agung Made Yoga Mahaputra⁴

^{1,2,3,4} Electrical Engineering Department, Politeknik Negeri Bali, Bali, Indonesia
swardika@pnb.ac.id

Abstract. Human body structures supported by the skeletal system are susceptible to osteoporotic vertebral compression fractures due to aging. The most common fracture among the geriatric population is vertebral fracture (25%), compared to hip fracture (24%). These symptoms, which largely stem from osteoporosis in post-menopausal women and men past their peak bone mass, carry significant health implications. Specifically, it can impair lung function, leading to a reduction in forced vital capacity of up to 9% and a higher risk of fatal pulmonary dysfunction. A diagnosis can be performed using invasive lateral radiography/X-rays, MRI, or bone scans to assess fracture severity. However, the lengthy and costly invasive diagnosis causes difficulties for the elderly. A new method of non-invasive diagnosis for human body structure is introduced. These new methods utilize pose-detection models. The 360-degree pose photos of the elderly are used to generate a 3D reality of elderly body models. This study results show the modified BlazePose detection successfully estimates 3D coordinates of key points of the body's skeleton. The BlazePose model achieved 97.2% accuracy at an average percent of correct points with 20% tolerance and RMSE of about 30.1 in millimeter units. Through 3D geometric analysis, results found degrees of misaligned of the shoulder, hips, and vertebral translations. The misaligned bone structures, indicated by more than 10 degrees, suggest a high likelihood of a spinal problem. This new method demonstrates an advancement in AI capabilities for medical applications.

Keywords: AI for Human Body Structure Diagnosis, Modified Blazepose Model, Non-invasive Osteoporotic Diagnosis, Pose Model Estimation with 3D Geometric Translations

1 Introduction

Nowadays, human body structure diagnosis uses invasive electromagnetic waves, i.e., X-rays, conventional Rontgen, computed tomography or CT scan, a computerized X-ray imaging, or magnetic resonance imaging (MRI) is a medical imaging technique that uses magnetic fields and radio waves to produce detailed images of the body's organs and tissues (Alsoof et al., 2022). Unlike CT scans or X-rays, MRIs do not use radiation. CT scans and MRI have been integrated with computer-aided diagnostic systems (CAD) (Seol et al., 2022). Furthermore, artificial intelligence (AI) has been integrated with these, enhancing accuracy and efficiency through image analysis and predictive modeling (Canoni-Meynet et al., 2022) However, the high cost of instrument access

© The Author(s) 2025

A. A. N. G. Sapteka et al. (eds.), *Proceedings of the International Conference on Sustainable Green Tourism Applied Science - Engineering Applied Science 2025 (ICOSTAS-EAS 2025)*, Advances in Engineering Research 280,

https://doi.org/10.2991/978-94-6463-878-3_8

can be challenging, especially for the elderly in large population countries. Today, everyone is aware of the advanced capabilities of AI, as demonstrated by the introduction of large language models, such as ChatGPT (Zhang & Shao, 2024). AI, through computer vision, has the potential to predict human body joints or landmarks using pose detection models.

This study leverages pose-detection models, AI computer vision techniques that identify human body landmarks (e.g., shoulder, hip) via regular RGB camera to reconstruct the spine in vertebral fracture osteoporosis (VFO) patients, thereby assisting the diagnostic process. With this technology, this manuscript introduces a novel method for non-invasive diagnosis of the human body structure. This manuscript presents a brief method, results, and discusses the diagnosis of human body structure, e.g., vertebral fracture osteoporosis, using pose detection models. Due to limited resources, this manuscript does not cover all standards or metrics benchmarks, e.g., accuracy, precision (Santiary et al., 2024; Swardika & Santiary, 2024). This study incorporates, in technical detail, i.e., how the dataset will be collected due to mobility limitations of elderly patients (Sumarwoto et al., 2023), how to improve existing pose models to be more specific for VFO patients by adding detailed spinal structure landmarks (Arya & Maji, 2024; Bazarevsky et al., 2020), how to reconstruct the spine of VCF patients in 3D perspectives that are obtained from landmark pose coordinates, and what analysis to be performed to predict, i.e., diagnostic degrees of spinal deviation.

2 Methodology

The methodology of this study begins with creating a specific platform that can automatically or manually rotate cameras surrounding elderly or VFO patients by 360 degrees or in full-body view. Because camera angles for datasets require consistency, the platform uses an adjustable camera holder shaft. As subjects that have been older, they cannot step and stand for longer. The platform must be sufficiently low, and quick action is required to process datasets. The camera is equipped with an ultra-wide lens of 0.5x and is sufficiently light. As the subject stands over the platform, the operator manually rotates the holder shaft and shoots many times. The following Figure 1 shows the platform and methods for collecting datasets.



Figure 1. The Platform and Methods for Collecting Datasets

Google Media Pipe’s human body pose detection model, powered by BlazePose, offers high-fidelity 3D pose tracking (Bazarevsky et al., 2020). It infers 33 landmarks and performs body segmentation in real-time across various platforms (mobile, desktop, web). The system uses two models: one for human detection (bounding box) and another for landmark identification within that box. Optimized with convolutional neural networks like MobileNetV2 and leveraging GHUM for 3D pose estimation (Grishchenko et al., 2022), it ensures efficient performance. Three model bundles are available, with input shapes of $224 \times 224 \times 3$ for the pose detector and $256 \times 256 \times 3$ for the pose landmarker. The following Table 1 shows details of model bundles.

Table 1. Google Media Pipe Pose Model Bundles

Model bundle	Format File	Complexity	Size (kB)	Data type	Model Cards	Versions
Lite	.task	1	5,643	float 16	info	Latest
Full	.task	2	9,178	float 16	info	Latest
Heavy	.task	3	29,946	float 16	info	Latest

The previously mentioned media Pipe pose models are provided as .task files, a format specifically for TensorFlow Lite models within media Pipe. These files package trained machine learning models for media tasks, such as object detection. There’s a trade-off between file size, accuracy, and detection speed: “Lite” bundles are smaller and faster but less accurate, while “Heavy” bundles are larger and slower but more accurate. All models output 33 3D landmarks. For this research, however, the focus is on four specific landmarks: left shoulder (11), right shoulder (12), left hip (23), and right hip (24), along with the midpoint of these four. Each landmark’s output includes its 3D coordinates (x, y, z) and visibility, provided in both normalized and world coordinates. The detection results for these four landmarks, their connecting lines, and three midpoints are visualized in a 3D XYZ plane using world coordinates. The following Figure 2 shows the modified BlazePose model for this study.

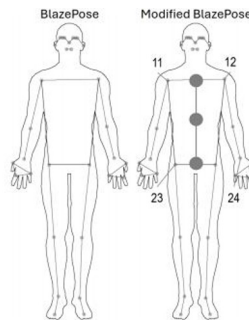


Figure 2. Modified BlazePose

The author’s earlier deep learning research focused on detecting objects in botanical and mental health contexts (Santiary et al., 2024; Swardika & Santiary, 2024). Now, this study ventures into creating pose detection models. This requires the author to learn

how to identify 3D pose landmarks using methods such as coordinate regression and heatmap approaches. Locating human body landmarks is a complex problem in visual computing. The figure illustrates this process, showing how an elderly person’s photo becomes a posing landmark. This involves using a powerful Convolutional Neural Network (CNN). Once an elderly person is detected, the system extracts the necessary landmarks by applying various techniques, including those from Spatial Transformer Networks (STN) and Feed-Forward Networks (FFN), and utilizing encode-decode steps to predict landmarks accurately (Grishchenko et al., 2022).

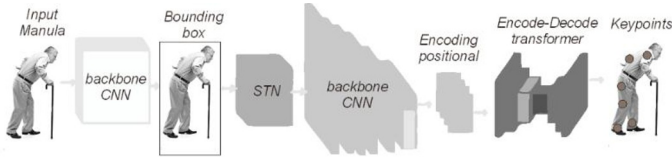


Figure 3. Stages in the Development of Pose Landmark Recognition Models

While the Blaze Pose models in Figure 3 above accurately identify 33 joint points, their 2D output doesn’t directly aid in diagnosing VFO, which is the focus of our research. To achieve this, the next crucial stage, detailed in the upcoming figure, involves converting these pose models into a 3D representation using four different photographic angles. This 3D reconstruction is crucial because it enables us to detect deviations from normal alignment, thereby facilitating the VFO diagnosis that this research aims to achieve. As a new model for human pose detection, the BlazePose uses the Percent of Correct Points at 20% tolerance (PCK@0.2) as an accuracy metric. This metric measures the 2D Euclidean error of a correctly detected key point that is less than 20% of the person’s torso size. The BlazePose model achieved 97.2% accuracy at an average PCK@0.2 on the human-annotated dataset. For the Root Mean Square Error or RMSE of measurement of the differences between the predicted and ground-truth key point locations, particularly in 3D pose estimation, the BlazePose model achieved a median RMSE of 30.1 in millimeter units.

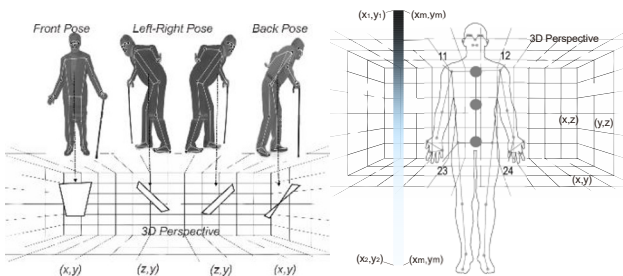


Figure 4. Reconstruction of Pose Results into a 3D Perspective and Analysis

To understand how to interpret datasets for analysis, emphasis to the use of degrees of deviation of geometric translations of the shoulder, hips, and vertebral lines as illustrated on the following Figure 4 above. For that, datasets are equipped completely

with a perpendicular reference line. This reference line is made by a bold band that is perpendicular to the bottom with gravity weight (a tripod with a red band in Figure 1). The diagnostic process involves a program that directly computes discrepancies in angles. Specifically, it calculates the difference between the vertical reference line (ϕ_R) and the backbone line (ϕ_B). Additionally, it determines the angular difference between the horizontal reference line (ϕ_{RH}), derived by rotating the vertical reference line 90 degrees, and both the shoulder line (ϕ_S) and the hip line (ϕ_H). In case of angles of reference line not detected automatically, the vertical reference lines are manually marked by mouse pointing (x_n, y_n points that are marked as green dots on the photo). Those points provide an average slope (m) result of the perpendicular reference line in degrees (ϕ) as computed with equation 1 below.

$$m = \tan \phi = 1/n \sum (y_2 - y_1) / (x_2 - x_1) \quad (1)$$

$$\phi = \tan^{-1} m$$

3 Results and Discussion

3.1 Result

The participants (subjects of VFO) in this study were elderly individuals, generally 50 years of age and older. A comprehensive collection of 24 datasets was compiled from 12 subjects, with each subject contributing two separate datasets. Within each dataset, 35 full-body photographs were captured from various angles. The following Figure 5 illustrates front-view examples from these 24 datasets, representing the study's VCF subjects.



Figure 5. Subjects of VFO Participants

The modified BlazePose heavy bundle model was selected to emphasize the highest accuracy output rather than FPS real-time high-speed processing. Only four landmarks pose than focused on (number: name and value: axis coordinates), i.e., PoseLandmark(i).name and PoseLandmark(i).value.x, y, z, and visibility. Where number and name are 11-12, left-right shoulder 23-24, left-right hip. The spinal or backbone landmarks are estimated from these. The result is a point landmark and a line connection over it.

As seen in Figure 2, there are seven landmark points in focus of this study, namely shoulder-right (S-R), shoulder-left (S-L), hip-right (H-R), hip-left (H-L), backbone-top (B-T), backbone-middle (B-M), and backbone-bottom (B-B). So, the confusion matrix for this study is presented in the following Table 2 because pose detection by means is not a standard classification or object detection when a correct prediction is not marked as their class's annotation or their names. However, marked as how close or far of geometric point is estimated within a 20% tolerance. Overall, the test results of this study show that all ground-truth key points are close to their landmark coordinates. With this result, the calculations of mean average precision (mAP) obtain 100%.

Table 2. The Confusion Matrix for This Study

		ground-truth key points						
		S-R	S-L	H-R	H-L	B-T	B-M	B-B
predicted key points	S-R	close						
	S-L	close						
	H-R	close						
	H-L	close						
	B-T	close						
	B-M	close						
	B-B	close						

The following Figure 6 shows the result of the landmark of subject VFO, their connection, and a plot in 3D perspective.

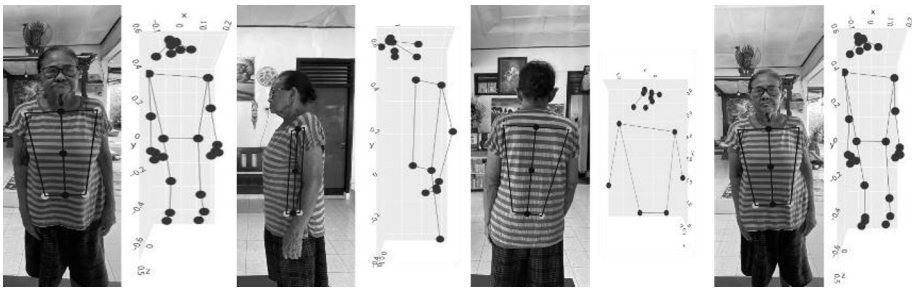


Figure 6. Landmarks Pose and 3D Perspectives Results

A subject VFO that had been more than 70 years old, easily recognized, had a spinal misalignment. The result of the figure above shows some spinal misalignment as a subject of VFO. However, units that can be used as a standard for measurement have not yet been published. This study uses degrees of deviation from a reference line as a unit of measurement. The following Table 3 shows the results of subject VFO with degrees of deviation from a reference line as a unit of measurement as diagnostics. The name of the subject is anonymous. In the Table, s = shoulder, h = hip, and b = backbone. Since a standard diagnostic degree scale hasn't been established yet, this research uses the following scale (based on subject NNS, an 82-year-old female): With an angle of inclination ≥ 10 degrees ($^{\circ}$), the likelihood of a high spinal problem (H) is indicated; others are marked as Low/Mid (L/M).

Table 3. Results of the Subject of VCF Diagnostics

Name of Subject VCF*	Diagnosis results (o) [s-h-b]			Scale
GAH(1) 59 Male	9.834692079858328	9.431805444726077	9.431805444726077	L/M
GAH(2) 59 Male	9.521221930661618	8.542306934346465	9.542306934346465	L/M
NNP(1) 57 Female	6.924240169222992	7.150746145746957	6.150746145746957	L/M
NNP (2) 57 Female	6.987067428058881	6.440768387576188	6.440768387576188	L/M
KA(1) 55 Female	6.237096414870248	9.05887852601105	9.05887852601105	L/M
KA(2) 55 Female	6.878356669581283	9.37500046937288	8.37500046937288	L/M
KO(1) 53 Female	5.423840129233237	6.124714881738166	6.124714881738166	L/M
KO(2) 53 Female	5.123840129233237	5.824714881738166	6.124714881738166	L/M
NNS(1) 82 Female	11.86476944752226	15.17531304113678	15.17531304113678	H
NNS(2) 82 Female	10.56476944752226	14.27531304113678	15.57531304113678	H
NWW(1) 79 Female	10.07046278195476	11.17329363657861	11.17329363657861	H
NWW(2) 79 Female	11.17046278195476	10.27329363657861	12.37329363657861	H
NWR(1) 55 Female	4.423060404355884	8.203514606368026	8.203514606368026	L/M
NWR(2) 55 Female	5.123060404355884	7.303514606368026	7.603514606368026	L/M
IWB(1) 57 Male	8.889339774724021	6.457296697828722	6.457296697828722	L/M
IWB(2) 57 Male	4.721297362764497	7.655040655235950	7.655040655235950	L/M
IKS(1) 58 Male	7.155699418461182	9.267012839502245	9.267012839502245	L/M
IKS(1) 58 Male	7.827105566967181	9.452640497299262	9.452640497299263	L/M
NKR(1) 63 Female	11.01046278195476	10.43329363657861	11.25329363657861	H
NKR(2) 63 Female	10.07046278195476	11.17329363657861	11.17329363657861	H

*(anonym)/Age/Sex

3.2 Discussion

The study of using non-invasive methods to predict human body structure is relatively new. As stated before, human body structure-related issues are internal to the body. Although AI technology has been integrated into these issues, it can only be implemented after the data acquisition phase. Charters et al. show cone-beam computed tomography (CBCT) and/or planar kV or MV projection imaging automated detection of vertebral body misalignments (Charters et al., 2024). Luximon et al. show detection of off-by-one vertebral-body misalignments using computed tomography (CT) images

and setup CBCT images (Luximon et al., 2022); another researcher also relies on invasive instruments (Alotaibi et al., 2022; D'Antoni et al., 2022). However, this study presents explainable results, even though they haven't yet been compared to other findings or established standards.

4 Conclusion

This study presents a method for diagnosing shifts in human body structure resulting from aging, birth defects, or accidents, particularly when bone fragility compromises spinal integrity. The approach utilizes pose detection models to pinpoint shoulder and waist landmarks, and then applies geometric principles to estimate the 3D coordinates of the spinal center. Diagnosis involves comparing the slopes of the shoulder, waist, and spinal lines against a reference, with the slope difference indicating the deviation from normal. A diagnostic scale is created using elderly subjects with visible bone fragility as a baseline for comparison. For improved scientific accuracy, the researchers suggest collaborating with physiologists to refine measurement criteria and incorporate relevant physiological research.

Acknowledgment

Thanks to Politeknik Negeri Bali for funding this study. This study is under contract number SP.DIPA-139.03.2.693476/2025 Rev. 03 date 07 March 2025.

References

- Alotaibi, G., Awawdeh, M., Farook, F. F., Aljohani, M., Aldhafiri, R. M., & Aldhoayan, M. (2022). Artificial intelligence (AI) diagnostic tools: Utilizing a convolutional neural network (CNN) to assess periodontal bone level radiographically—a retrospective study. *BMC Oral Health*, 22(1), 399.
- Alsoof, D., Anderson, G., McDonald, C. L., Basques, B., Kuris, E., & Daniels, A. H. (2022). Diagnosis and management of vertebral compression fracture. *The American Journal of Medicine*, 135(7), 815–821.
- Arya, V., & Maji, S. (2024). Enhancing human pose estimation: A data-driven approach with mediapipe BlazePose and feature engineering analysis. *2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT)*, 1–6.
- Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). *BlazePose: On-device real-time body pose tracking* (No. arXiv:2006.10204). arXiv.
- Canoni-Meynet, L., Verdout, P., Danner, A., Calame, P., & Aubry, S. (2022). Added value of an artificial intelligence solution for fracture detection in the radiologist's daily trauma emergencies workflow. *Diagnostic and Interventional Imaging*, 103(12), 594–600.
- Charters, J. A., Luximon, D., Petragallo, R., Neylon, J., Low, D. A., & Lamb, J. M. (2024). Automated detection of vertebral body misalignments in orthogonal kV and MV guided radiotherapy: Application to a comprehensive retrospective dataset. *Biomedical Physics & Engineering Express*, 10(2), 025039.

- D'Antoni, F., Russo, F., Ambrosio, L., Bacco, L., Vollero, L., Vadalà, G., Merone, M., Papalia, R., & Denaro, V. (2022). Artificial intelligence and computer-aided diagnosis in chronic low back pain: A systematic review. *International Journal of Environmental Research and Public Health*, 19(10).
- Grishchenko, I., Bazarevsky, V., Zanzir, A., Bazavan, E. G., Zanzir, M., Yee, R., Raveendran, K., Zhdanovich, M., Grundmann, M., & Sminchisescu, C. (2022). *BlazePose GHUM holistic: Real-time 3D human landmarks and pose estimation* (No. arXiv:2206.11678). arXiv.
- Luximon, D. C., Ritter, T., Fields, E., Neylon, J., Petragallo, R., Abdulkadir, Y., Charters, J., Low, D. A., & Lamb, J. M. (2022). Development and interinstitutional validation of an automatic vertebral-body misalignment error detector for cone-beam CT-guided radiotherapy. *Medical Physics*, 49(10), 6410–6423.
- Santiary, P. A. W., Swardika, I. K., Dewi, D. A. I. C., & Sugirianta, I. B. K. (2024). Intra-class deep learning object detection on an embedded computer system. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 13(1), 430-439.
- Seol, Y. J., Kim, Y. J., Kim, Y. S., Cheon, Y. W., & Kim, K. G. (2022). A study on 3D deep learning-based automatic diagnosis of nasal fractures. *Sensors*, 22(2).
- Sumarwoto, T., Hartanto, D., & Utomo, P. (2023). Managing fractures in geriatrics: Current approaches and update. *JKKI: Jurnal Kedokteran dan Kesehatan Indonesia*, 69–81.
- Swardika, I. K., & Santiary, P. A. W. (2024). Detection of vague object signatures on deep learning surveillance devices. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 13(3), 3262-3272.
- Zhang, H., & Shao, H. (2024, March 1). *Exploring the latest applications of OpenAI and ChatGPT: An in-depth survey*. EBSCOhost.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

