Innovations

Advanced AI-Driven Spine Posture Detection Using CNN and Mediapipe for Accurate Postural Analysis and Correction

[1]Mr. M. Priyadharshan; [2]Ms. Bhoomiya S K*; [3]Mr. Kirubhakaran S; [4]Ms. Abinaya R; [5]Ms. Subhashini S G
Assistant Professor, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India
Department of Computer Science and Engineering,
Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India

Corresponding Author: **Bhoomiya S K**

Abstract: *Sedentary lifestyles, particularly among office workers, have increased the prevalence of spinal health issues stemming from poor posture and misalignment. These issues are often linked to chronic pain, reduced mobility, and diminished productivity, highlighting the urgent need for effective solutions. Traditional ergonomic tools, such as adjustable chairs and desks, lack the capability to provide real-time feedback essential for sustained posture correction. This study presents an AI-powered spine posture detection system leveraging Convolutional Neural Networks (CNNs) and Media Pipe's pose estimation framework. By analyzing spinal alignment in real time, the system identifies deviations and delivers immediate corrective feedback to encourage healthier postural habits. The integration of CNNs' deep learning capabilities with Media Pipe's efficient pose-tracking ensures accurate and scalable posture analysis across diverse workplace environments. Extensive testing demonstrates the system's reliability in detecting misalignments under varying conditions, promoting improved posture and mitigating risks associated with prolonged spinal strain. Limitations, such as reduced accuracy in unconventional sitting positions, are discussed, alongside future directions like adaptive algorithms and personalized feedback.This research highlights the transformative potential of AI in addressing ergonomic challenges, offering a practical solution to improve spinal health, enhance workplace productivity, and foster long-term musculoskeletal well-being.*

Keywords: *Body posture detection, Chronic back pain, Deep learning, Ergonomic health, Health Care*

1. Introduction:

The increasing prevalence of sedentary lifestyles, particularly among professionals in technology-oriented industries, has led to a surge in spine-related health issues. Extended periods of sitting in non-ergonomic positions are a primary cause of spinal strain, persistent back pain, and musculoskeletal disorders, all of which have detrimental effects on health and workplace productivity.NIH research reveals prolonged inactivity and non-ergonomic environments correlate with chronic spine problems, with 80% of desk-based workers experiencing back issues. Prolonged poor posture can worsen conditions such as kyphosis, lordosis, and scoliosis, often resulting in limited mobility, nerve compression, and chronic pain.Methods such as MRI, CT scans, and X-rays, while accurate for evaluating musculoskeletal configuration, are impractical for regular clinical use due to logistical challenges, costs, and safety concerns, especially the risk of increased radiation exposure[1].

This study addresses these challenges by proposing a real-time spine posture detection system powered by deep learning and computer vision technologies. The system integrates Convolutional Neural Networks (CNNs) with the MediaPipe framework to continuously monitor body posture and provide instant feedback on any deviations from optimal alignment. This approach builds on advancements in ergonomic health technologies, which emphasize the importance of posture-related health risks.

By combining the robust pattern recognition abilities of CNNs with MediaPipe's precise skeletal tracking capabilities, the proposed system ensures accurate posture detection and classification. Such tools aim to promote healthier sitting habits and reduce the long-term health implications of poor spinal alignment**.** Artificial Intelligence significantly influences various aspects of daily life, making it evident that future lifestyles for most individuals will involve increased sedentary behaviour. As a result, it is essential to investigate and implement strategies to minimise the negative impacts of this inevitable sedentary lifestyle[5]. This study underscores the potential of deep learning solutions in enhancing musculoskeletal health and improving overall well-being in environments characterized by prolonged sedentary behaviour.

Abbreviations	**Explanation**
NIH	National Institutes of Health
CNN	Convolutional Neural Network
DL	Deep Learning
AI	Artificial Intelligence
RMSE	Root Mean Square Error
ReLU	Rectified Linear Unit
FPS	Frames Per Second

Table 1. Abbreviation Table

2. Methodologies

It integrates Convolutional Neural Networks (CNNs) with Media Pipe to enable efficient skeletal pose estimation and classification. The methodology ensures precise spinal alignment analysis, delivering real-time feedback to encourage better postural habits.

2.1. Convolutional Neural Networks (CNN) Framework

2.1.1. Architecture

The backbone of the posture detection system is a CNN framework optimized for image classification. The architecture consists of multiple convolutional layers designed to extract hierarchical features from the input images. These layers are interspersed with pooling layers, which reduce spatial dimensions and computational complexity while preserving critical features as in Figure 1. Fully connected layers follow, aggregating the extracted features for binary classification into "good" or "bad" posture. The design ensures that even subtle differences in posture are captured effectively, building on techniques inspired by previous studies [4],[5].

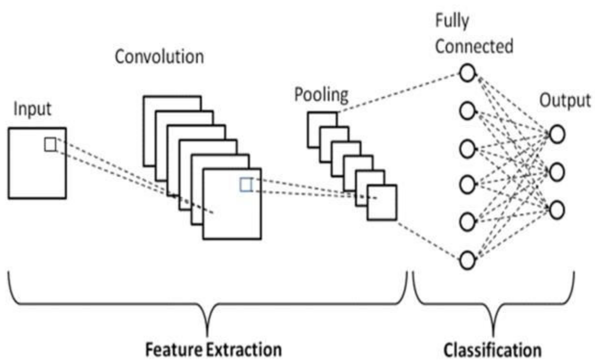

Figure 1. Convolutional Neural Network

2.1.2. Activation and Regularization

ReLU (Rectified Linear Unit) activation functions are incorporated to introduce non-linearity, allowing the model to learn complex patterns in the data. The data overfitting was set to 0.7, and the maximum number of iterations in trainingepochswas set to 100.To combat overfitting, dropout layers are strategically placed within the network. These layers randomly deactivate a fraction of neurons during training, ensuring the model does not become overly reliant on specific features. Binary cross-entropy loss is employed to measure classification errors, ensuring robust learning. Finally, the output layer uses a sigmoid activation function, producing probabilistic outputs to classify posture into binary categories. These methodologies, supported by prior research [9],[10] underpin the CNN's high performance and reliability.

2.2. Data Preprocessing and Dataset

2.2.1. Normalization

Normalization is a fundamental preprocessing step in the provided code, applied to the dataset before training the Convolutional Neural Network (CNN). This process involves scaling the pixel values of the images from their original range of 0 to 255 to a standard range of 0.0 to 1.0 by dividing each pixel value by 255.0. This adjustment ensures that the data is compatible with the activation functions used in the network, such as ReLU and sigmoid, which perform more effectively within this range. Moreover, normalization improves the stability of the training process by preventing large gradients during backpropagation, leading to more stable weight updates and faster convergence. It also ensures that all input features have a similar scale, which helps avoid bias towards high-intensity pixel values. By maintaining consistency across the training and validation datasets, normalization reduces the risk of overfitting and enhances the interpretability of the learned features. This preprocessing step is simple yet crucial for ensuring the model processes the data efficiently and achieves optimal performance.

2.2.2. Dataset

The dataset for spine posture detection plays a critical role in training the convolutional neural network (CNN) and ensuring its accuracy in identifying good and bad posture. It is organized into two primary categories: "good posture" and "bad posture," with each category containing images stored in separate directories. The "good posture" directory includes images of individuals maintaining correct posture, such as a straight spine and aligned shoulders, while the "bad posture" directory contains images of individuals exhibiting improper postures, such as slouched shoulders or a bent spine. These images are manually curated or sourced from real-life and simulated scenarios to represent a wide range of postural variations. Each image is preprocessed before being fed into the model. The preprocessing steps include resizing the images to 100×100 pixels to match the CNN’s input dimensions, normalizing the pixel values to a range of [0, 1], and assigning binary labels—1 for good posture and 0 for bad posture. These steps standardize the dataset, allowing the model to learn consistently across varying input dimensions and pixel intensity levels.

The dataset is then split into training and validation sets, with 80% of the images used for training and 20% for validation. For instance, if the dataset contains 1,000 images, 800 are allocated for training, and 200 are reserved for validation. This split ensures the model learns from a substantial amount of data while being evaluated on unseen data to monitor its generalization ability. The training set includes diverse examples of postures, enabling the model to identify complex patterns and key visual features associated with good and bad postures. These features include the alignment of shoulders and hips, the inclination of the spine, and body orientation. During the validation phase, the model is tested on its ability to generalize these learned features to new images, ensuring it performs well on data it has not encountered during training. A balanced dataset is crucial to prevent bias toward one class. For example, if the dataset contains 600 images of good posture and 400 of bad posture, the model is trained on an equal representation of both categories, promoting fairness in predictions.

The CNN analyzes each image's features to determine the posture status. Images in the "good posture" category often have visually distinct patterns, such as symmetrical alignment, which the model learns to associate with a higher probability of good posture (e.g., >0.7). Conversely, images in the "bad posture" category exhibit misalignments or curvatures, leading the model to assign a lower probability (e.g., ≤0.7). This probability threshold allows real-time predictions, where frames from a live video feed are processed, resized, and passed through the CNN to predict posture status. The model's output, combined with the computed spine inclination angle, enables detailed feedback, such as displaying "Good Posture" or "Bad Posture" on the screen. Additionally, this system can issue warnings for sustained poor posture, reinforcing the importance of corrective action. A robust dataset with well-prepared images is key to the model's success, ensuring accurate and reliable detection across diverse postural scenarios.

2.3. Real-Time Pose Estimation with Media Pipe

MediaPipe plays a pivotal role in extracting skeletal landmarks essential for spinal alignment assessment. Key landmarks, such as the shoulders and hips, are tracked to determine the posture. Using geometric computations, the system calculates angles between these landmarks, quantifying deviations from an ideal alignment. This method leverages techniques described in earlier studies [8], ensuring precision and reliability.

To detect misalignment, the system defines a threshold for angular deviations. If the deviation exceeds this predefined limit, the system triggers alerts, prompting users to adjust their posture. MediaPipe's lightweight framework ensures efficient landmark extraction, enabling real-time performance without compromising accuracy. The integration of pose estimation with CNN-based classification enhances the system's utility in diverse environments, making it a robust solution for posture monitoring.

2.3.1. Technologies and Tools

Python serves as the primary programming language, leveraging its rich ecosystem of libraries. Tensor Flow and Keras are used for building and training the CNN. Media Pipe provides the pose estimation capabilities necessary for real-time skeletal tracking. Supporting libraries, such as NumPy and Matplotlib, are employed for data manipulation and visualization, respectively. The integration of these tools creates a cohesive framework that supports the system's end-to-end functionality, reflecting best practices in the field.

3. Graph Representation for Visual Analysis

3.1 Model Accuracy and Signs of Over fitting: The model accuracy plot, represented on the left, consists of two key performance indicators: training and validation accuracy. The training accuracy, shown by a blue line, consistently increases throughout the epochs, suggesting that the model is effectively learning from the training data. By the final epoch, the training accuracy reaches around 0.9, reflecting high classification accuracy for the training dataset.

However, the validation accuracy, represented by an orange line, exhibits significant fluctuations, hovering between 0.4 and 0.7. Unlike the training accuracy's steady increase, this inconsistency in validation accuracy indicates potential overfitting. Overfitting occurs when the model learns to perform well on the training data but fails to generalize effectively to new, unseen data. Addressing this challenge is crucial, as it directly impacts the practical application of the model for real-world posture detection. Techniques such as dropout layers, data augmentation, and more varied training data can help improve generalization.

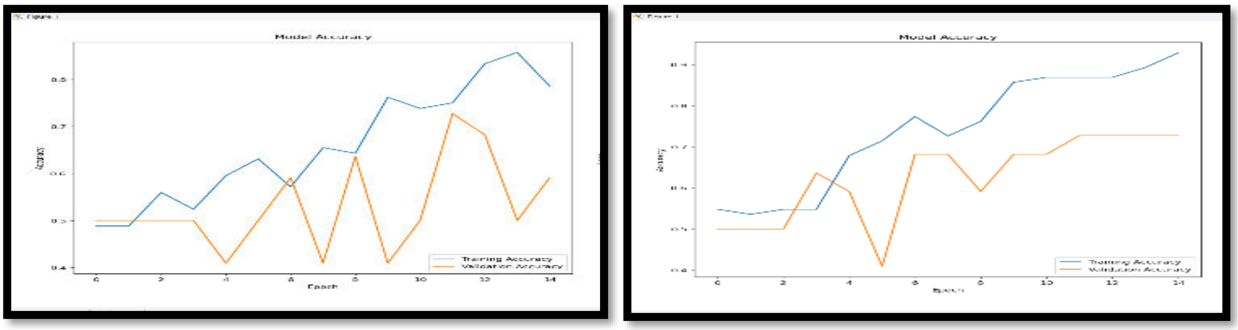

Figure 2. Model Accuracy

3.2 Model Loss Analysis The model loss plot on the right reinforces observations made from the accuracy graph. The blue line represents training loss, which consistently decreases over time, indicating that the model is successfully minimizing prediction errors on the training data. This downward trend is a positive indicator, showing that the model is learning effectively.

In contrast, the validation loss, depicted by the orange line, fluctuates and sometimes increases in specific epochs. This erratic behaviour is another sign of overfitting, where the model performs well during training but exhibits inconsistent performance when evaluated on validation data. This variability underscores the need for additional regularization or data diversification to ensure that the model can handle new inputs more reliably.

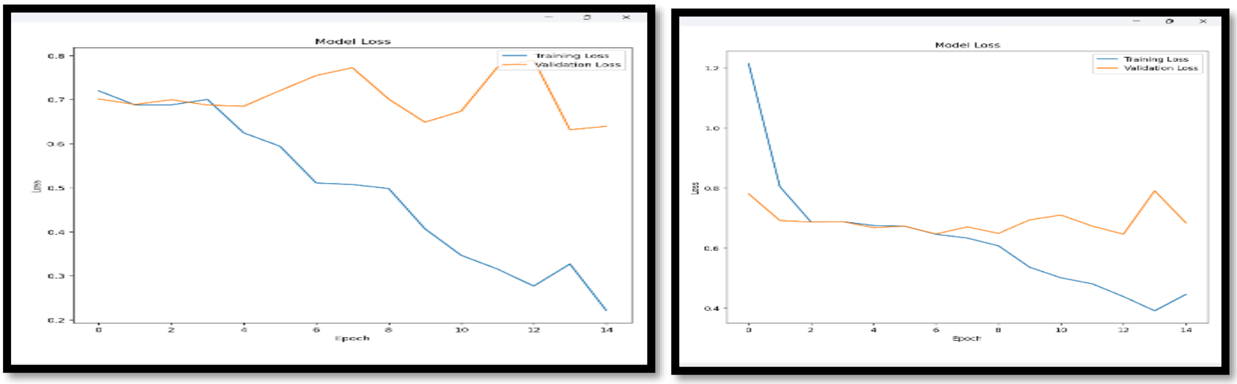

Figure 3. Model Loss

4. Angle Detection

The angle detection in this project is implemented to provide a quantitative assessment of posture through mathematical computations. The primary function,findangle, calculates the angle between defined body points, specifically the shoulder and hip, to determine the inclination of the spine. This calculation is based on the dot product formula for vectors, which is pivotal in deriving the cosine of the angle between two points.

The mathematical process involves computing the Euclidean distance between points using the Pythagorean theorem, followed by the application of the arccosine function. This returns the angle in radians, which is then converted into degrees. The formula used can be represented as

$$\theta = \arccos\left(\frac{(y2-y1).(-y1)}{\sqrt{(x2-x1)^2+(y2-y1)^2.y1}}\right)(1)$$

where θ is the angle between the vertical axis and the line connecting the two landmarks.

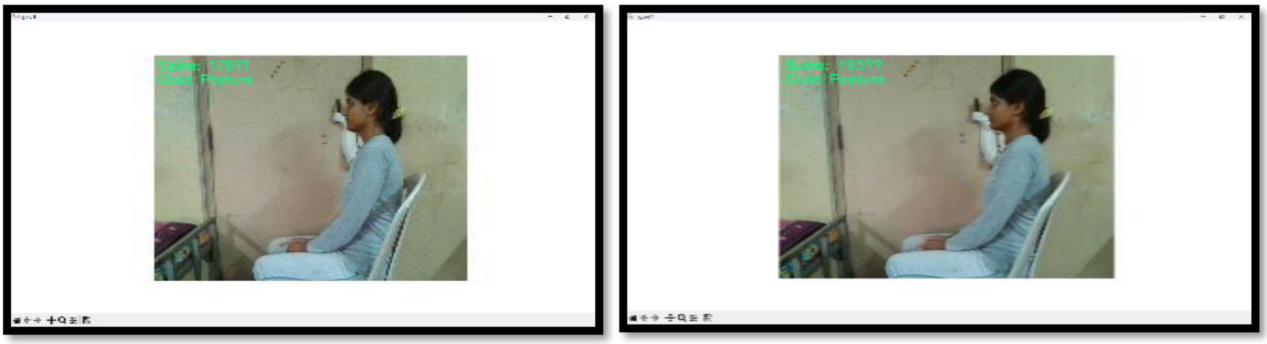

Figure 4. Good Posture

Accurate angle calculation is vital for distinguishing between subtle changes in posture. For instance, an angle closer to 90° typically indicates good posture, while significant deviations may signal slumping or leaning. The angle detection is integrated into the real-time analysis, triggering warnings if the inclination exceeds a predefined threshold over an extended period.

The function's reliability is enhanced by considering edge cases, such as partial visibility of body parts or rapid movements that may skew angle detection. To mitigate such issues, the model incorporates error-handling mechanisms and uses temporal smoothing, averaging angles over a set number of frames to provide consistent posture evaluations.

Figure 5. Bad Posture

These angle measurements are visualized alongside posture status, giving users immediate feedback on their body alignment. The graphical overlay displaying the spine angle ensures that users are informed of their real-time posture condition, making corrective action more intuitive.

Aspect	CNN	MediaPipe
Purpose	Primary used for image classification, object detection, and other deep learning tasks involving vision.	A framework for building pipelines to process multimodal data (video, audio, etc.), with a focus on real-time computer vision tasks.
Core Technology	Deep learning model, uses convolutional layers to automatically extract features from raw image data.	Framework that uses pre-built models for tasks like hand tracking, face detection, and pose estimation. Relies on pretrained models and efficient pipelines.
Model Type	A type of neural network that is trained end-to-end for specific tasks.	Provides a set of pre-trained models and efficient processing pipelines designed for real-time inference.
Customization	Highly customizable; users can design and train models specific to their needs using datasets.	Limited customization compared to CNNs; focuses more on providing easy-to-use, pre-built models and pipelines.
Real-Time Processing	Can Process image or video frames in real-time if optimized well, but usually requires powerful hardware for real-time performance.	Optimized for real-time performance across various devices (e.g., mobile, edge devices), with high efficiency built into the framework.
Implementation	Requires substantial expertise in deep learning, data preprocessing, and model training.	Simplifies implementation with ready-to-use pipelines, making it easier for developers to integrate complex computer vision tasks into applications.

Table – 2. Difference between Convolutional Neural Network and Mediapipe

5. Mathematical Concept Used

The mathematical underpinnings of this project are critical for understanding how posture is evaluated and classified. The project leverages fundamental geometry and trigonometry to calculate distances and angles. The function *findDistance* employs the Pythagorean theorem to measure the straight-line distance between two points in a 2D space, represented as

$$d = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \qquad (2)$$

This calculation forms the basis for determining the spatial relationships between landmarks.

The *findAngle* function is pivotal for assessing posture. By calculating the angle formed between the shoulder and hip, the function determines whether the spine is upright or inclined. This angle is derived using the dot product of vectors and the arccosine function, ensuring precision in angle measurement. Conversion from radians to degrees is performed using the formula

$$degrees = \theta \times (\frac{180}{\pi})(3)$$

This step ensures that the system can assess whether the spine is aligned or tilted based on the calculated angle.

5.1 Normalization

During the training phase of a **Convolutional Neural Network (CNN)**, image normalization is applied to scale pixel values to a uniform range, typically between 0 and 1. This process ensures consistency across input data and speeds up training by stabilizing the model.

$$normalized\ pixel\ value = \frac{pixel\ value}{255} \qquad (4)$$

Where:
The normalized pixel value results in a range between 0 and 1.

6. Conclusion

This project demonstrated the successful creation of a CNN-based spine posture detection system, integrating deep learning and MediaPipe for effective real-time analysis. By following a structured design, the project incorporated pre-trained pose estimation models and custom-built CNNs for accurate classification of spinal posture. The CNN architecture was designed to strike a balance between complexity and generalization, with convolutional layers, max pooling, dropout for regularization, and a sigmoid activation function optimized for binary classification.

Although the model showed strong training performance, variability in validation accuracy highlighted potential overfitting, indicating that future improvements should focus on greater data diversity and model optimization. The system was able to process webcam feeds to identify posture deviations and issue alerts if poor posture persisted, fostering proactive user adjustments for better spinal health. Future enhancements could include expanding the dataset, refining data augmentation techniques, and adding multi-class classification for detecting a wider range of posture variations. This project underscores the value of AI-driven posture monitoring systems for preventative health, promoting continued innovation in adaptive models that are resilient and ready for real-world applications.

References

1. *Odesola D F., Kulon J., Verghese S., Partlow A., Gibson C, Smart Sensing Chairs for Sitting Posture Detection, Classification, and Monitoring: A Comprehensive Review, Sensors 2024,24(9), 2940.*
2. *V. Sreevani., B. Sathwik Reddy., K. Nithin., K. Harsha Vardhan., Kahtan A. Mohammed., Uma Reddy, Sorabh Lakhanpal., Ravi Kalra, Advanced interdisciplinary approaches for bad posture detection using computer vision and IoT,vol 507, ICFTEST-2024,E3S Web of Conferences 507, 01045,2024.*
3. *W. Armstrong, J. Borg, M. Krizack, Guidelines on the Provision of Manual Wheelchairs in Less-Resourced Settings,World Health Organization, 2008, 2022,www.who.int.*
4. *Mujtaba Hussain Jaffery, Muhammad Adil Ashraf, Ahmad Almogren, Hafiza Mahnoor Asim,Jehangir Arshad ,Javed Khan, Ateeq Ur Rehman , Seada Hussen, FSR-Based Smart System for Detection of Wheelchair Sitting Postures Using Machine Learning Algorithms and Techniques, vol.1, Article ID 1901058, 2022.*
5. *Aminosharieh Najafi, T., Abramo, A.; Kyamakya, K.; Affanni, A. Development of a Smart Chair Sensors System and Classification of Sitting Postures with Deep Learning Algorithms, Sensors 2022, 22, 5585.*
6. *X. Ran, C. Wang, Y. Xiao, X. Gao, Z. Zhu, and A. Bin Chen, "Portable sitting posture monitoring system based on a pressure sensor array and machine learning," Sensors and Actuators A: Physical, vol. 924, article 112900, p. 4247, 2021.*
7. *W. Cun, R. Mo, J. Chu., "Sitting posture detection and recognition of aircraft passengers using machine learning," Artificial Intelligence for Engineering Design, Analysis and Manufacturing, vol. 35, no. 3, pp. 284–294, 2021.*
8. *Q. Wan, H. Zhao, J. Li, and P. Xu, "Hip positioning and sitting posture recognition based on human sitting pressure image," Sensors, vol. 21, no. 2, p. 426, 2021.*
9. *Chaitanya Kumar A., V G Sridhar, Design and Analytics of Smart Posture Monitoring System,IOP Publishing, J. Phys.: Conf. Ser. 2115 012048,2021.*
10. *S. Liaqat, K. Dashtipour, K. Arshad, K. Assaleh, N. Ramzan, "A hybrid posture detection framework: integrating machine learning and deep neural networks," In IEEE Sensors Journal, vol. 21, no. 7, pp. 9515–9522, 2021.*