biophysica

MDPI

Article

DeepSarcAE: A Deep Autoencoder Framework for Learning Gait Dynamics in the Detection of Sarcopenia

Muthamil Balakrishnan [1], Janardanan Kumar [2], Jaison Jacob Mathunny [1], Varshini Karthik [1] and Ashok Kumar Devaraj [1,*]

[1] Department of Biomedical Engineering, SRM Institute of Science and Technology—Kattankulathur, Chengalpattu 603203, India; mb9162@srmist.edu.in (M.B.); jaisjac@gmail.com (J.J.M.)
[2] Department of General Medicine, SRM Medical College Hospital and Research Centre, SRM Institute of Science and Technology—Kattankulathur, Chengalpattu 603203, India
* Correspondence: ashokd@srmist.edu.in

Abstract

Sarcopenia is a degenerative musculoskeletal condition recognised as the age-related decline in skeletal muscle mass, strength, and function. Traditional diagnostic methods are limited by cost, accessibility, and subjectivity. This study aimed to develop a non-invasive, AI-driven, video-based framework for early Sarcopenia detection using functional movement analysis. Participants with and without Sarcopenia were recorded performing functional movements such as level walking, stair climbing, and ramp walking. Ten representative frames were extracted from each participant, resulting in 300 images (150 Sarcopenic, 150 non-Sarcopenic) utilised for the study. The DeepSarcAE model is an integrated framework of an autoencoder and a CNN-based classifier. Its performance was benchmarked against pretrained architectures such as EfficientNet, ResNet, MobileNet, Inception, VGG16 and four custom CNN models. Evaluation metrics such as sensitivity, specificity, precision, negative predictive value (NPV), accuracy, and AUC were used to analyse the models. DeepSarcAE outperformed all other models, attaining 100% sensitivity, 83.33% specificity, 85.71% precision, 100% NPV, 91.67% accuracy, and an AUC of 0.96. VGG16 and MobileNet followed the performance of DeepSarcAE closely, while the Inception network exhibited the weakest results due to poor generalisation. TheDeepSarcAE framework offers a scalable, cost-effective, and non-invasive approach for Sarcopenia screening from the input gait image frames. Its promising preliminary performance highlights the potential of deep learning in early diagnosis and clinical decision support in preventive healthcare.

Keywords: biomechanical image analysis; gait analysis; video-based diagnosis; functional movement; non-invasive screening

Received: 9 February 2026
Revised: 3 March 2026
Accepted: 9 March 2026
Published: 16 March 2026

1. Introduction

Sarcopenia is a degenerative musculoskeletal condition recognised as the age-related decline in skeletal muscle mass, strength, and function. Due to its clinical relevance in geriatric health, it is officially classified as a disease under the ICD-10 (M62.84) [1]. This disorder considerably leads to frailty, and an increased risk of falls, hospitalisation, and mortality in the elderly population. The pathophysiology of Sarcopenia involves several factors such as neuromuscular degeneration, hormonal imbalances, chronic inflammation, nutritional deficiencies, etc. [2].

The global prevalence of Sarcopenia varies from 8% to 36% in those under 60 years and ranges from about 10% to 27% in those aged 60 years and above. A meta-analysis

indicated a pooled prevalence of 10% among community-dwelling older persons, with elevated rates noted among hospitalised and institutionalised populations [3]. The burden of Sarcopenia is more significant in Asia because of the demographic changes and lifestyle factors. India has a prevalence of approximately 43.6% in Sarcopenia [4], with regional variations resulting from dietary and socioeconomic differences. South India exhibits a significantly elevated prevalence of approximately 14.2% among older populations, as indicated by community-based studies conducted in Tamil Nadu [5].

In order to standardise the diagnosis of Sarcopenia, expert groups have suggested several clinical guidelines, such as the European Working Group on Sarcopenia in Older People (EWGSOP2). It describes a step-by-step process to evaluate muscle strength, muscle quantity or quality, and physical performance in the diagnosis of Sarcopenia [6]. The Asian Working Group for Sarcopenia (AWGS 2019) adapted these criteria in order to fit regional populations, implementing cut-off values for handgrip strength, gait speed, and appendicular skeletal muscle mass [7]. Despite these improvements, conventional diagnostic methods, such as dual-energy X-ray absorptiometry (DEXA), bioelectrical impedance analysis (BIA), and computed tomography (CT), still face significant challenges. DEXA and BIA are influenced by the hydration status and variations between different machines. CT and MRI, on the other hand, are considered among the gold standards but are not easy to conduct in primary care settings [8].

CT-based assessment of the skeletal muscle index (SMI) at the third lumbar vertebra (L3) has demonstrated a robust correlation with total body muscle mass. However, manual segmentation with tools like Slice-O-Matic is time-consuming and it requires radiological knowledge, which makes it impractical for large-scale screening, especially in places where resources are limited [9].

Timely diagnosis of Sarcopenia is essential in order to initiate treatments that can lower healthcare expenses, enhance the quality of life for those affected, and ultimately delay the progression of the disease. Recent studies have demonstrated a significant interest in investigating the aetiology of Sarcopenia and in developing precise measurement techniques [10]. The level of physical activity, extended periods of inactivity, cardiovascular endurance, and muscular strength are essential determinants in evaluating the risk of Sarcopenia. Therefore, in rural areas where expert physicians are not available, an automated method for diagnosing and managing Sarcopenia based on clinical and kinetic parameters is required.

Artificial intelligence (AI), particularly image analysis based on deep learning, has emerged as a promising approach in addressing these limitations. Recent research has shown that AI can predict the risk of Sarcopenia by integrating the features from images with clinical and biomechanical features [11–13]. Such technologies hold potential for the early detection of Sarcopenia and personalised intervention, especially in underserved regions like rural South India, where access to specialist care is limited.

Dual-energy X-ray absorptiometry (DXA), computed tomography (CT), and magnetic resonance imaging (MRI) are traditional methods utilised to estimate muscle mass. These techniques are reliable in practice but are time-consuming and very expensive, and not always available in primary care or community settings. This has prompted increasing interest in artificial intelligence (AI)-driven, non-invasive approaches to functional movement assessment as a scalable solution for Sarcopenia screening.

Recent research implies that gait and posture analysis are critical windows into Sarcopenia. Kim et al. [14] developed a classification model which integrates smart insole data with video-based pose estimation. The studies show that gait deviations, particularly hip angle asymmetries, distinguish Sarcopenic from non-Sarcopenic subjects with a high classification accuracy. In a similar study, Perez-Lasierra et al. [15] reviewed gait assess-

ments based on inertial measurement units (IMUs) and concluded that the spatiotemporal gait features analysed through machine learning classifiers offer strong predictive utility. However, differences in sensor protocols make it hard to standardise these assessments.

Wearable sensor approaches are highly popular in the diagnosis of Sarcopenia. For instance, research utilising IMUs placed on the lower back or ankles has derived features from various gait phases, using algorithms such as Support Vector Machines (SVMs), Random Forests (RFs), and Multilayer Perceptrons (MLPs), achieving reliable classification performance [16]. More recent multimodal frameworks are designed to combine smart insoles, IMUs and computer vision with sensor fusion, improving the detection of Sarcopenia and other related cognitive decline [17]. However, depending on specialised devices makes it harder to operate in large-scale or resource-limited settings.

Research suggests video-based methodologies are a promising non-invasive alternative for diagnosing Sarcopenia. Juan et al. designed a sit-to-stand video analysis app to demonstrate the robust detection of Sarcopenia. The study is carried out among community-dwelling older adults by leveraging pose estimation algorithms to capture movement deficits without requiring wearable devices [18]. Despite these advances in the diagnosis of Sarcopenia, several limitations exist in traditional methods. Most studies focus exclusively on level walking while overlooking more demanding functional activities such as stair ascent and descent or slope negotiation. This may reveal subtler impairments in strength, coordination, and balance. Previous research involves validation of AI-based gait analysis as a feasible tool in the diagnosis of Sarcopenia. These studies still have challenges in movement diversity, scalability, and generalizability. The study overcomes these challenges by leveraging deep learning techniques with biomechanical analysis which are derived from videos across multiple functional contexts in the diagnosis of Sarcopenia.

This work develops a novel AI-based approach in Sarcopenia diagnosis using video-derived biomechanical analysis across multiple functional movement phases such as parallel walking, stair climbing, and slope negotiation. The work relies solely on the sagittal video capture and hence eliminates the need for wearable sensors or imaging modalities. This greatly simplifies data acquisition and enhances the scalability and feasibility of Sarcopenia screening in both clinical and community settings. This process enables scalable and non-invasive screening of Sarcopenia with the aid of deep learning techniques. Thus, this framework addresses the limitations of existing diagnostic approaches by providing a practical tool to detect Sarcopenia in clinical and community settings.

The primary aim of this study is (i) to record functional movement videos of participants from Sarcopenic and non-Sarcopenic groups during three movement phases such as parallel walking, stair climbing, and slope negotiation; (ii) to implement and compare multiple deep learning architectures (ResNet50, VGG16, MobileNetV2, InceptionNet, EfficientNet) for feature extraction and the classification of Sarcopenia; (iii) to develop and validate a non-invasive, AI-based diagnostic framework for the diagnosis of Sarcopenia by analysing video-derived biomechanical patterns during multiple functional movement tasks. The study utilises deep learning architectures to precisely categorise individuals as Sarcopenic or non-Sarcopenic, facilitating a scalable and early detection in clinical and community environments.

2. Material and Methods

The methodology designs an automated system for diagnosing Sarcopenia based on gait analysis from the video recordings. The process involves data acquisition, video-to-image conversion, preprocessing, feature extraction through pretrained and custom networks, and final classification into Sarcopenia and non-Sarcopenia. Figure 1 describes the overall block diagram of the work. Initially, the gait videos were captured from the

participants in walking tasks across three distinct terrains, such as parallel walking, stair climbing, and ramp ascent and descent. By utilising transfer learning, deep visual features were extracted from the selected images and classified using a set of pretrained convolutional neural networks (CNNs) such as ResNet50 [19], VGG16 [20], InceptionNet [21], MobileNetV2 [22], and EfficientNet [23]. These five architectures were selected to represent a diverse range of CNN design paradigms: ResNet50 for deep residual learning, VGG16 for uniform deep convolutional feature extraction, InceptionNet for multi-scale feature aggregation, MobileNetV2 for lightweight and computationally efficient inference, and EfficientNet for balanced model scaling. This selection enabled a comprehensive evaluation of architectural trade-offs in terms of depth, complexity, and efficiency for the small-scale biomedical image classification task. In addition, a custom-designed DeepSarcAE (Deep Sarcopenia Autoencoder) network was developed to learn compact, latent gait representations that encode subtle biomechanical deviations. Thus, this end-to-end approach facilitates non-invasive, efficient, and scalable screening of Sarcopenia from a video input rather than complex imaging or wearable sensor systems.

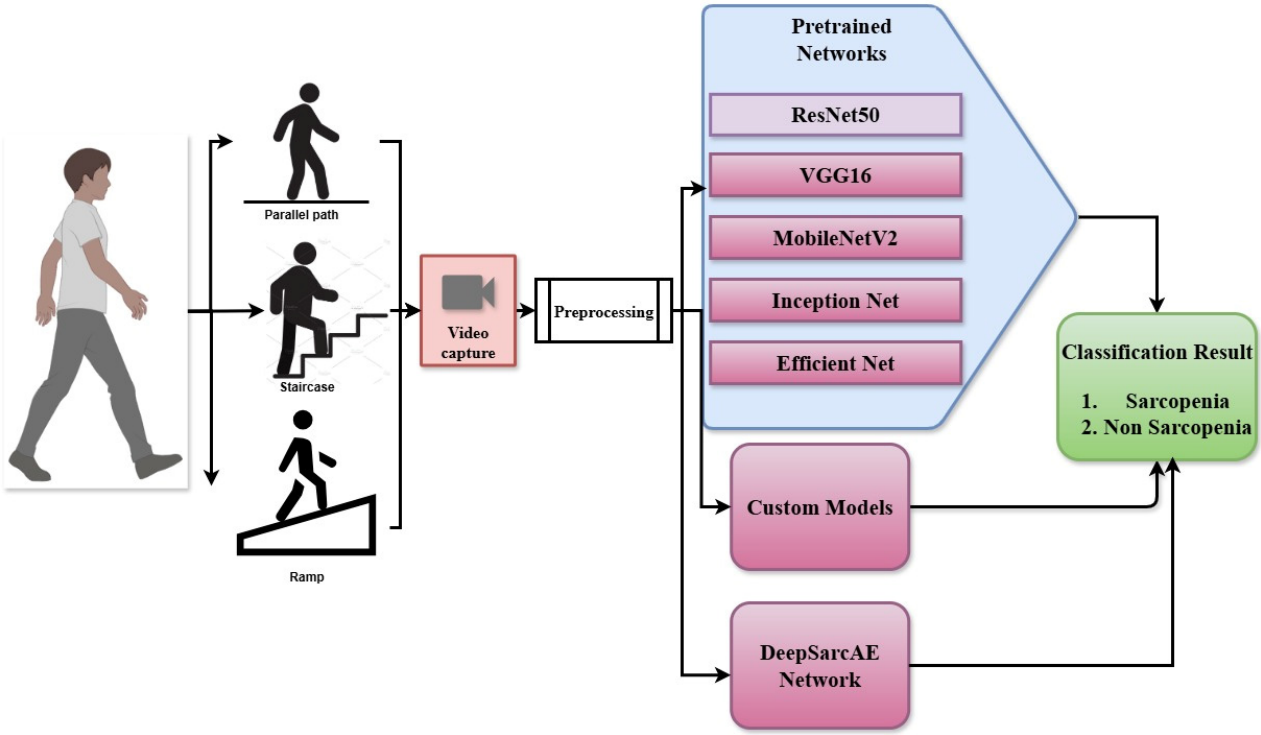

Figure 1. Overview of the Sarcopenia detection framework. Gait videos are recorded while participants walk on a parallel path, climb stairs, and traverse ramps. Extracted video frames undergo preprocessing before feature extraction using pretrained CNNs (ResNet50, VGG16, InceptionNet, MobileNetV2, EfficientNet), custom models, and the DeepSarcAE network, and are classified into Sarcopenic and non-Sarcopenic categories.

2.1. Data Collection

The study was carried out in the Department of Physical Medicine and Rehabilitation at SRM Medical College Hospital and Research Centre, Kattankulathur, India from June 2024 to January 2025. The study obtained ethical permission with clearance number 3075/IEC/2022 from the Human Ethical Committee of SRM Hospital & Medical College & Research Centre, Kattankulathur. About 30 individuals were classified into two groups, Sarcopenic and non-Sarcopenic, based on physician assessment. Participants with known musculoskeletal disorders (such as osteoarthritis or joint replacement), neurological conditions (such as neuropathies, stroke, or Parkinson's disease), or any other pathology

that could independently affect gait patterns were excluded from the study to minimise confounding biomechanical influences. The research employed a smartphone (Oppo F19 Pro) (OPPO Electronics Corp., Dongguan, Guangdong, China) to capture videos for angle assessment. The videos were captured with the smartphone's rear camera, preserving a recording resolution of 1920 × 1080 p and a frame rate of 30 fps. The camera was positioned 2 m beyond the centre of the path and calibrated to their hip level in both the sagittal and transverse planes.

Both groups were instructed to walk at their preferred pace along all three pathways equipped with a guardrail. The stepped terrain comprised four steps, each with a consistent elevation of 16 cm, resulting in a total vertical height of 64 cm. This configuration was established to evaluate joint kinematics during ascent and descent, and to emphasise the enhanced demands imposed on the knee and ankle joints during vertical displacement. The sloped pathway was an inclined plane that was 182 cm long and 63 cm high at its top, making the angle of inclination around 20 degrees. This moderate inclination was designed to evaluate the functional performance of lower limb joints, like the knees and ankles, as they are constantly raised. Finally, the parallel path or level walking surface was a flat, 230 cm long corridor, which allowed subjects to walk at a self-selected pace. Ten images were extracted from the videos of each subject, and hence a total of 300 images with 150 Sarcopenic and 150 non-Sarcopenic were included in the study. Experienced physicians clinically evaluated all participants in the dataset based on the European Working Group on Sarcopenia in Older People (EWGSOP2). The labels provided by physicians served as annotations for supervised learning, ensuring clinical validity and reliability in the training and evaluation of the model.

2.2. Video Capture and Preprocessing

The work involves capturing videos in order to record the natural gait movements representing the functional daily activities. The participants were recorded while performing three distinct locomotor tasks, such as walking on a parallel path, ascending and descending a staircase, and walking on an inclined ramp. All participants followed a fixed sequential order of tasks: parallel path walking was performed first, followed by stair climbing, and finally ramp walking. This standardised order was maintained across all participants to ensure consistency in data acquisition and to minimise variability arising from task sequencing. Each participant's gait video was processed to extract ten representative frames that captured critical gait events. From the ramp walking sequence, three frames were selected for uphill motion (heel strike, toe off, foot flat) and three for downhill motion (heel strike, toe off, foot flat). Two frames were chosen from the stair-climbing task (heel strike and toe off), and two additional frames were taken from the parallel path sequence (heel strike and toe off).

The preprocessing pipeline is designed with several steps to standardise and to enhance the image data. All the extracted frames were first resized to a fixed dimension suitable for deep learning input, and then pixel intensity normalisation was performed to minimise the lighting discrepancies across sessions. In order to isolate the silhouette and to minimise environmental interference, background noise was reduced from the images using Otsu's thresholding-based segmentation combined with morphological filtering (erosion and dilation) to isolate the participant's silhouette from the background. Minor adjustments in contrast and sharpness were carried out to emphasise the body contours and joint articulation which improves the visual clarity of the movement patterns.

2.3. Feature Extraction and Classification Using Pretrained Networks

The preprocessed images comprising 300 representative gait frames were utilised for deep learning-based feature extraction and classification. The study utilises pretrained CNN architectures, such as ResNet50, VGG16, MobileNetV2, InceptionNet, and EfficientNet, through transfer learning to detect Sarcopenia. These models were trained on the ImageNet database and can be utilised to provide low- and mid-level visual features in Sarcopenia diagnosis by freezing the initial layers with weights and altering the final classification layer.

2.4. Custom Models

The study designs four customised CNN models to diagnose Sarcopenia with distinct architectural variations in order to evaluate their effectiveness in the classification. Custom Model 1 is designed with a lightweight autoencoder integrated with a shallow CNN classification network. Image reconstruction is carried out using the encoder–decoder structure of the autoencoder through two encoding and decoding stages, each with convolutional layers of 32 filters. This enables basic feature compression and recovery. The CNN framework of this model consists of five convolutional layers (increasing with 8 to 128 filters), followed by three dense layers (512, 256, and 128 neurons). This model utilises Exponential Linear Unit (ELU) activation in order to promote faster convergence and to mitigate vanishing gradient issues. Custom Model 2 incorporates a deeper and more complex encoder–decoder structure with 512 convolutional filters and several down-sampling layers. The architecture predominantly contains ReLU activation function in order to ensure non-linear transformations with efficient gradient propagation. When compared with Custom Model 1, the deeper design of this model provides greater representational power in complex imaging tasks. On the other hand, the computational complexity of the model increases with the risk of overfitting.

Custom Model 3 employs GELU as the activation function [24] in its layers in order to achieve smoother gradient flow with better regularisation. Although the encoder–decoder framework is similar to Model 1, the convolutional layers of Model 2 have more filters and are deeper. The classification framework contains multiple GELU-activated dense layers that enable nuanced feature interactions. Custom Model 4 is the CNN framework from the DeepSarcAE network without the autoencoder. It is designed to study the impact of the autoencoder in the model. Overall, the architectures of the designed custom models progresses from shallow to deep architectures, from ELU to GELU activations, and from simple to complex encoders, reflecting a systematic exploration of depth, activation choice, and representation learning.

2.5. DeepSarcAE: Autoencoder-Based Feature Extraction and Classification Framework

The study designs the DeepSarcAE framework with a deep learning approach, as shown in Figure 2, to predict Sarcopenia by extracting high-level biomechanical and postural features from the images. This network overcomes the limitations faced by the pretrained models, such as ResNet or Inception, by providing a lightweight, task-specific, and computationally efficient alternative that can be trained from scratch, even with limited medical image data. The framework combines the strengths of convolutional autoencoders for representation learning with a custom CNN classifier optimised for binary classification of the input image into Sarcopenic and non-Sarcopenic participants [25].

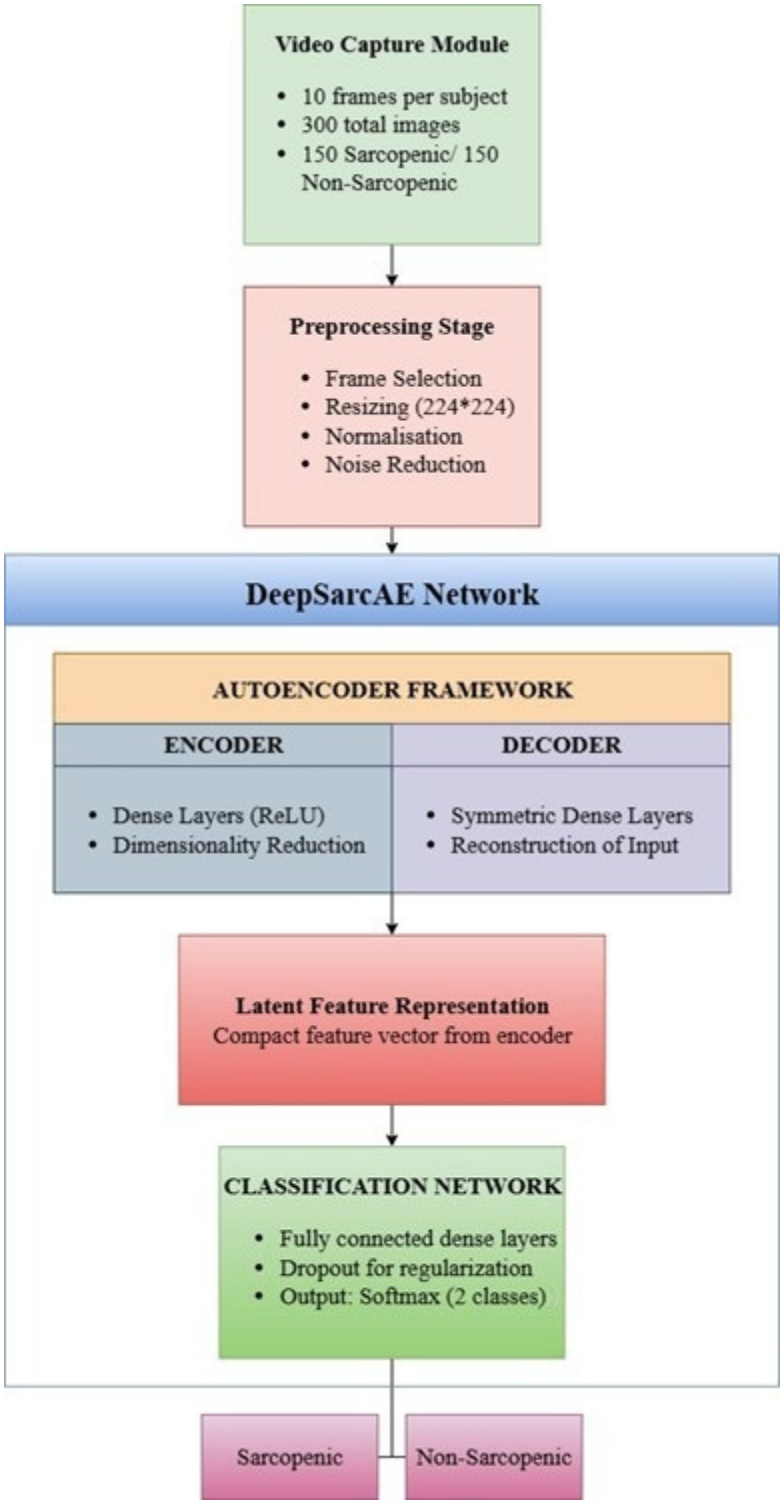

Figure 2. Overall system workflow for video-based gait analysis and DeepSarcAE classification.

2.5.1. Architecture Design and Structural Overview

The architecture of DeepSarcAE, as shown in Figure 3, is made up of two primary modules, namely the encoder and the decoder. Both of these modules play complementary roles in compressing and then reconstructing the image representations. RGB gait images with an input size of 128 × 128 × 3 are given as input to the model in order to ensure a balance between computational tractability and the preservation of fine-grained biomechanical details. The encoder captures hierarchical visual features, while the decoder reconstructs the original image from the learned latent space.

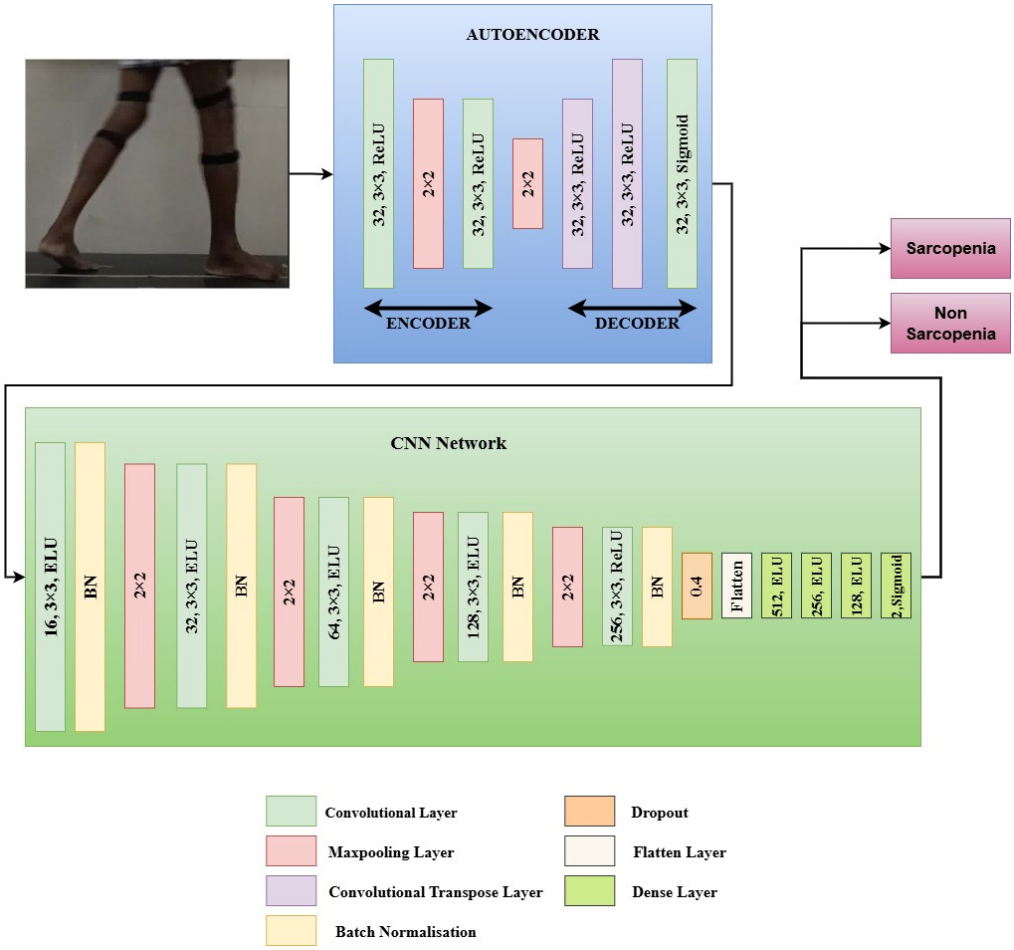

Figure 3. Architecture of the DeepSarcAE model for automated Sarcopenia classification.

2.5.2. Encoder Design

The encoder is designed with a series of convolutional and pooling layers in order to reduce the spatial dimensions and expand feature depth. The first convolutional layer consists of 32 filters with a kernel size of about (3 × 3). The ReLU activation function is used to maintain the spatial resolution through "same" padding. This layer is capable of detecting basic visual primitives such as edges, contours, and limb boundaries in the input image provided. This is followed by a MaxPooling layer with a (2 × 2) window to reduce the spatial dimensions by half and condense the information while retaining the dominant activations. This is followed by a second convolutional block, which consists of 32 filters, capturing the mid-level features such as limb orientation, angular relationships between joints, and texture variations in body posture. After the second pooling operation, the dimensions of the feature map are reduced to (32 × 32 × 32). This creates a compact representation that summarises both global and local spatial dependencies, which are relevant to gait characteristics of Sarcopenia.

2.5.3. Decoder Design

The structure of the decoder is similar to that of the encoder, but it replaces the convolution in pooling layers with convolution in transpose layers in order to upsample the generated feature maps back into the original image size. Two transposed convolutional layers, each with 32 filters and ReLU activations, are employed to reconstruct the spatial resolution progressively. The final reconstruction layer then employs a convolution layer with a sigmoid activation function in order to ensure that the pixel intensities lie within the range of [0, 1]. This configuration generates the reconstructed gait images which are

visually approximate in size to the input frames. The decoder is defined as a mapping $\hat{X} = g_\phi(Z)$, where g_ϕ reconstructs the input using learnable parameters ϕ. The training objective minimises the binary cross-entropy loss between the original image X and its reconstruction $\hat{X}$:

$$L(X, \hat{X}) = -\frac{1}{N}\sum_{i=1}^{N}[X_i \log(\hat{X}_i) + (1 - X_i)\log(1 - \hat{X}_i)]$$

This loss makes the model learn a compact and informative latent space which retains biomechanically relevant variations. The autoencoder was tuned using the Adam optimiser, with a learning rate of 0.001 and a batch size of about 128.

2.5.4. CNN Classifier for Sarcopenia Detection

The second stage of the DeepSarcAE network is a custom CNN classifier which is trained on the encoded features from the autoencoder. This framework interprets the latent spatial representations and then assigns class probabilities corresponding to Sarcopenic or non-Sarcopenic conditions. The classifier begins with convolutional layers with filter sizes of 16, 32, 64, 128, and 256, each followed by ELU (Exponential Linear Unit) activation functions [26]. These layers provide faster convergence and reduce the vanishing gradient problem. Batch Normalisation layers are utilised, which stabilise the activations and accelerate the training process. Progressive MaxPooling operations were included that reduce the spatial dimensions, and a Dropout layer (0.4 rate) randomly deactivates neurons during training to mitigate overfitting. Given that the model contains approximately one million trainable parameters while the training set consists of 240 images, the network operates in an overparameterised regime. To address the risk of overfitting, multiple regularisation strategies were employed: Dropout (rate of 0.4) was applied after convolutional blocks, Batch Normalisation was used to stabilise activations, and data augmentation techniques including random horizontal flipping, rotation (up to 15 degrees), and brightness adjustment were applied during training to artificially expand the effective dataset size.

The obtained feature maps are flattened into a one-dimensional vector and are passed through fully connected dense layers with 512, 256, and 128 neurons, with ELU activation functions. The final Dense layer uses sigmoid as activation in order to produce probabilities for the two classes. The model was finally optimised using Stochastic Gradient Descent (SGD) optimizer with a Nesterov momentum of learning rate = 0.01, decay = 1×10^{-6}, and momentum = 0.9.

3. Results

The current study captures the gait video frames and then classifies them using various pretrained and custom deep learning models. These models were assessed using five performance metrics, such as sensitivity, specificity, precision, negative predictive value (NPV), and AUC, as shown in Table 1. Derived diagnostic measures such as the false positive rate (FPR), false negative rate (FNR), F1-score, and overall accuracy are provided in Table 2. Figures 4–6 illustrate the receiver operating characteristic (ROC) curves obtained for pretrained, custom, and DeepSarcAE architectures, respectively, and Figure 7 shows the confusion matrix of the DeepSarcAE model.

Table 1 demonstrates that models VGG16, Custom Model 1, and DeepSarcAE have superior potential in the classification of Sarcopenia. Among the pretrained architectures, VGG16 achieved strong discriminative ability with high sensitivity (100%), specificity (80%), and the AUC (1.00). The DeepSarcAE model achieved results comparable to VGG16, with a sensitivity of about 100%, specificity of about 83.33%, and an AUC of about 0.96. These

results demonstrate its ability to generalise effectively to unseen test data. The effectiveness of the autoencoder-based feature learning pipeline is evident from the Custom Models 1 and 3 achieving AUC values of about 0.96 and 0.86, respectively.

Table 1. Performance metrics of pretrained, custom, and hybrid deep learning models for Sarcopenia classification based on gait video-derived images.

Model	Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)	AUC
EfficientNet	60.00	85.71	75.00	75.00	0.84
ResNet	66.67	83.33	80.00	71.43	0.86
MobileNet	100.00	77.78	60.00	100.00	0.83
Inception	50.00	33.33	42.86	40.00	0.46
VGG16	100.00	80.00	87.50	100.00	1.00
Custom Model 1	100.00	71.43	71.43	100.00	0.96
Custom Model 2	100.00	55.56	42.86	100.00	0.77
Custom Model 3	75.00	100.00	100.00	66.67	0.86
Custom Model 4	83.33	83.33	83.33	83.33	0.92
DeepSarcAE	100.00	83.33	85.71	100.00	0.96

Table 2. Derived diagnostic indices and overall accuracy of deep learning models for Sarcopenia detection.

Model	False Positive Rate (1 − Specificity) (%)	False Negative Rate (1 − Sensitivity) (%)	F1-Score (%)	Accuracy (%)
EfficientNet	14.29	40.00	66.67	75.00
ResNet	16.67	33.33	72.73	75.00
MobileNet	22.22	0.00	74.99	83.33
Inception	66.67	50.00	46.34	41.67
VGG16	20.00	0.00	93.33	91.67
Custom Model 1	28.57	0.00	83.33	83.33
Custom Model 2	44.44	0.00	60.00	66.67
Custom Model 3	0.00	25.00	85.71	83.33
Custom Model 4	16.67	16.67	83.33	83.33
DeepSarcAE	16.67	0.00	92.31	91.67

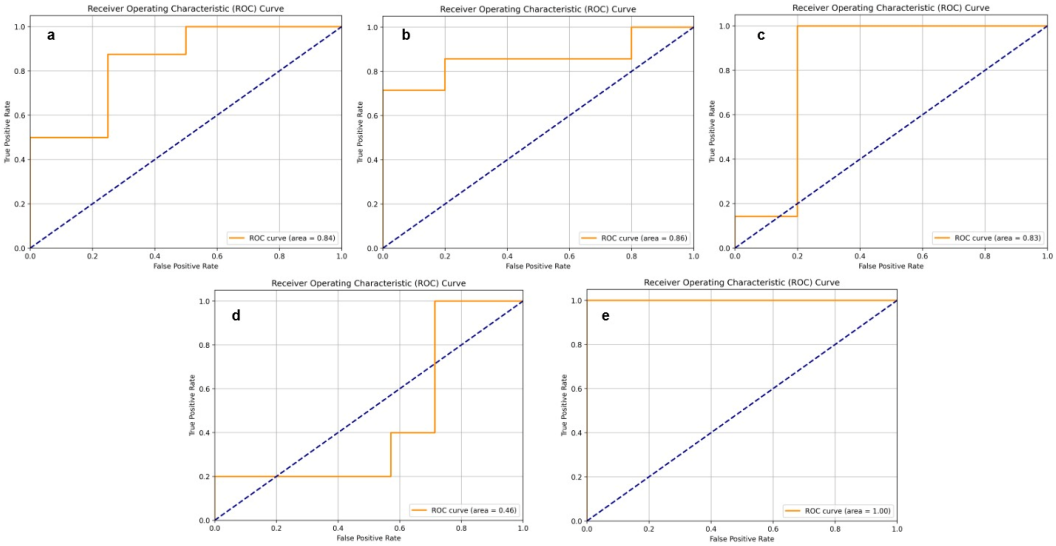

Figure 4. ROC curves of pretrained CNN models (EfficientNet, ResNet, MobileNet, Inception, and VGG16) for Sarcopenia classification. (**a**) EfficientNet; (**b**) ResNet50; (**c**) MobileNetV2; (**d**) InceptionNet; (**e**) VGG16.

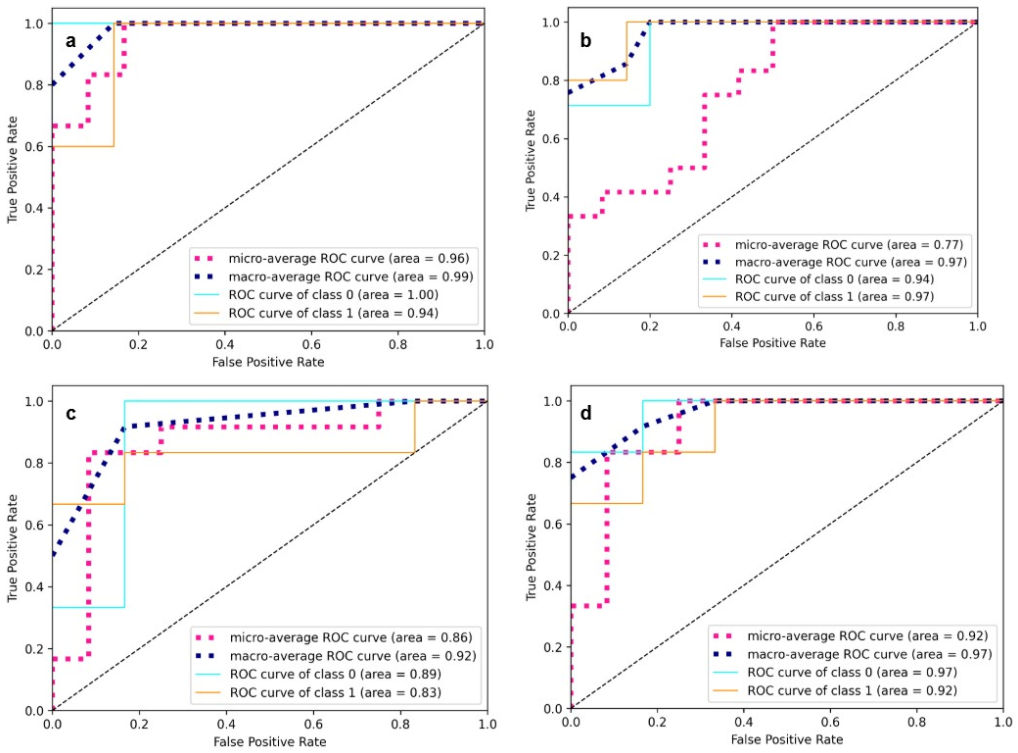

Figure 5. ROC curves of custom CNN architectures (Custom Models 1–4) trained on gait-derived biomechanical images. (**a**) Custom Model 1; (**b**) Custom Model 2; (**c**) Custom Model 3; (**d**) Custom Model 4.

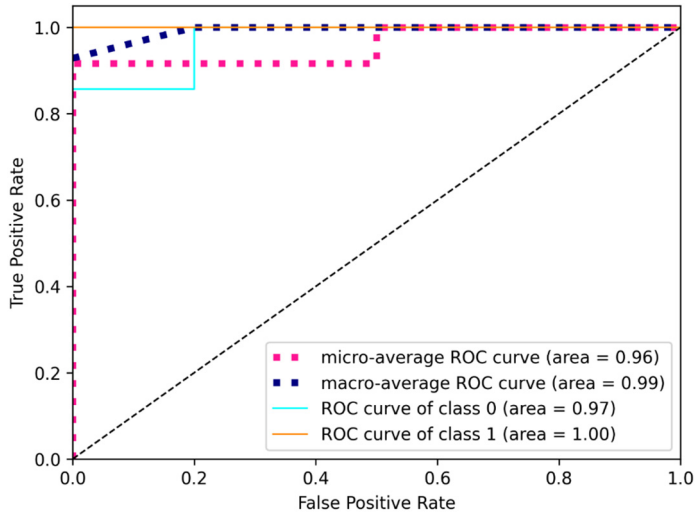

Figure 6. ROC curve of the DeepSarcAE hybrid model in classification of Sarcopenia.

Models such as MobileNet, VGG16, Custom Models 1 and 2, and DeepSarcAE had high sensitivity with a low false negative rate. This demonstrates their reliability in ensuring that no Sarcopenic subject was misclassified as non-Sarcopenic. The clinical significance of minimising false negatives paves the way for the early detection of Sarcopenia, which can enable a timely intervention. However, Custom Model 2 had a higher false positive rate of about 44.44%, indicating its tendency to overclassify non-Sarcopenic images.

As depicted in Figure 4, the ROC curve of the VGG16 network closely approached the top-left corner of the plot, illustrating the near-perfect discrimination between Sarcopenic and non-Sarcopenic images. MobileNet and ResNet also exhibited a stable ROC behaviour, with AUC values of about 0.83 and 0.86, respectively. The ROC curves for the custom models in Figure 5 emphasise the importance of architectural customisation in the small

dataset learning scenarios. Custom Model 1 demonstrated a curve shape closely aligned with that of VGG16, achieving an AUC of 0.96, which reflects its strong generalisation and balanced recall–precision profile. Custom Model 3 also achieved a high separability with an AUC of about 0.86, showing that its GELU activation and optimised feature extraction pipeline successfully reduced false positives. Custom Model 4 displayed an AUC of about 0.92. This indicates that a moderate architectural depth combined with GELU activation can produce consistent outcomes.

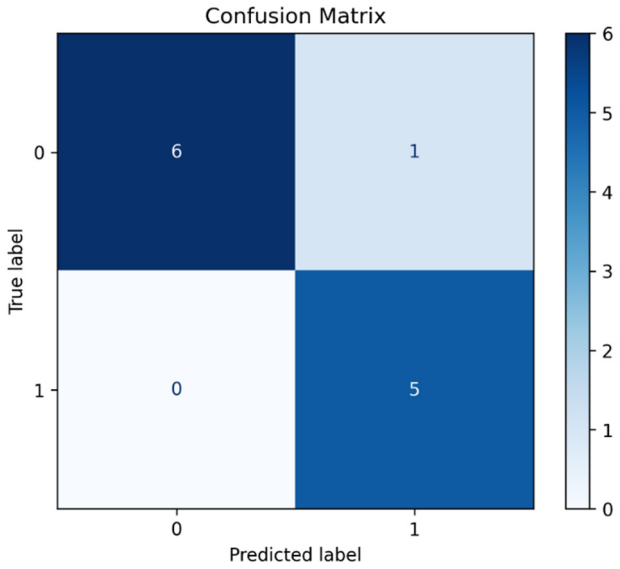

Figure 7. Confusion matrix of the DeepSarcAE model on the test dataset.

The DeepSarcAE model outperformed most of the architectures by integrating autoencoder-based feature extraction with a deep convolutional classifier. As seen in Figure 6, the ROC curve of DeepSarcAE is similar to that of VGG16, with an AUC of 0.96. This proves its strong classification stability and a balanced trade-off between sensitivity and specificity. It is important to note that the train/test split was performed at the subject level rather than the image level. All ten frames extracted from a given participant were assigned exclusively to either the training, validation, or test set, ensuring that no data leakage occurred between subsets. The dataset of 300 images was divided into training, validation, and test subsets using a tiered split. First, 80% (240 images) were used for model training. The remaining 20% (60 images) were reserved for validation and testing. Within this block, 20% (12 images) were allocated for final testing, while the remaining 48 images served as the validation set. Accordingly, the confusion matrix in Figure 7 reflects predictions on these 12 test samples (six Sarcopenic and six non-Sarcopenic). These results suggest that DeepSarcAE effectively captured the discriminative latent representations of Sarcopenia-associated gait and posture characteristics.

The ROC curves illustrated in Figures 4–6 prove the quantitative findings reported. VGG16 and DeepSarcAE resulted in near-ideal classification by demonstrating steep curve trajectories approaching the upper-left boundary. Custom Model 1, ResNet and MobileNet curves occupied the intermediate zone, confirming their balanced and slightly conservative discrimination thresholds, whereas InceptionNet traced a near-diagonal ROC curve reflecting minimal predictive power. It is seen that DeepSarcAE and VGG16 are the most effective architectures for predicting Sarcopenia from the biomechanical images. The DeepSarcAE framework thus represents a promising, interpretable, and computationally efficient approach for non-invasive Sarcopenia screening, though these results should be interpreted as preliminary given the limited sample size.

4. Discussion

The results of the work show the effectiveness of deep learning architectures in the automated classification of Sarcopenia using biomechanical images captured from gait videos. Although pretrained and custom convolutional models achieved promising accuracy in diagnosing Sarcopenia, the DeepSarcAE framework and the VGG16 network outperform all others by exhibiting the most balanced and robust performance. These results confirm that gait and posture analysis, when combined with CNNs and autoencoder-based feature extraction, can be used in the non-invasive diagnosis of Sarcopenia.

Among the pretrained networks, VGG16 had an outstanding discriminative ability by achieving a perfect sensitivity of 100% with a high specificity of about 80%, and the highest AUC of about 1. This suggests that the deep, uniform convolutional layers of the VGG16 network have effectively captured the subtle biomechanical markers of Sarcopenia from the sagittal-plane images. The Inception model, on the other hand, had poor performance, with an AUC of about 0.46. This indicates that the multi-branch architecture of InceptionNet failed to generalise effectively in a relatively small, homogeneous dataset. Even though EfficientNet and ResNet achieved a balanced specificity (83–85%) and AUCs above 0.84, their moderate sensitivity (60–66%) limited their clinical applicability since missing positive cases are critical in early Sarcopenia detection. Similarly, MobileNet achieved a perfect sensitivity of about 100% and zero false negatives. But its moderate precision (60%) causes more false positives (22.22%).

The custom CNNs yielded complementary insights into architectural optimisation for small biomedical datasets. Custom Model 1 achieved an AUC of about 0.96 and an accuracy of about 83.33%, comparable to that of the VGG16 network. Its encoder–decoder feature extraction stage provided a compact latent representation. This filters noise and retains discriminative kinematic features by providing a compact latent representation. Although Custom Model 3 achieved a slightly lower sensitivity of about 75%, it achieved a perfect specificity and precision of about 100%, emphasising its reliability in identifying non-Sarcopenic subjects without false positives. Custom Model 4 balanced all of these trade-offs by reaching both 83.33% sensitivity and specificity with an AUC of about 0.92. These results, integrated with the ROC in Figure 5, highlight that moderate depth, well-chosen activation functions, and balanced pooling layers are more significant for biomechanical classification than very deep architectures.

The DeepSarcAE framework achieved the most balanced and clinically relevant results, combining the advantages of unsupervised feature learning with supervised classification. The model eliminated the false negatives completely and hence maintained a low false positive rate of about 16.67%. Its ROC curve is closely paralleled with that of VGG16, confirming its comparable discriminative power in identifying Sarcopenia. The confusion matrix illustrates that DeepSarcAE accurately predicted the actual cases of Sarcopenia with reduced misclassification of the negative Sarcopenic subjects. The encoder–decoder structure of the autoencoder enables it to preserve the essential biomechanical signatures and discard the irrelevant visual noise. This process is proven to enhance class separability by allowing the classifier to focus on the subtle posture and gait variations induced by Sarcopenic muscle degeneration.

From the computational perspective, the high AUC values of VGG16, Custom Model 1, and DeepSarcAE (1.00, 0.96, and 0.96, respectively) are due to their structural simplicity and feature compression. This results in improved generalisation of the model under limited data conditions. The hybrid approach of the DeepSarcAE network improved interpretability and reduced computational cost in predicting Sarcopenia.

The study proves the significance of AI-assisted diagnostics in Sarcopenia detection. This video-based approach leverages movement patterns captured in everyday environ-

ments without the need for any specialised equipment such as DXA, CT, or bioimpedance analysis. Thus, the DeepSarcAE framework can be used as a cost-effective and non-invasive screening tool that could be deployed in primary care settings or telemedicine platforms.

The study has several limitations as follows. The primary limitation is the utilisation of a relatively small dataset with only 300 images, which are derived only from the sagittal-plane gait videos. This restricts the generalizability of the network across diverse populations. Data were collected under controlled conditions using the same device. Hence, they do not reflect real-world variability, such as lighting, camera angles, and surface differences. Also, the binary classification approach did not account for any intermediate Sarcopenia severity levels. This limits the clinical granularity. Additionally, the current study did not correlate the classification results with conventional imaging data such as CT or DXA-based muscle mass indices. Future work should integrate imaging-derived ground truth measures to further validate the model predictions against established diagnostic standards. Finally, while the DeepSarcAE model achieved high diagnostic performance, its interpretability remains limited. Hence, further explainable AI methods are required in order to identify which biomechanical cues will influence disease predictions.

Future research should include a much larger and more diverse dataset, obtained in a multi-view or 3D motion analysis, and temporal modelling using 3D CNNs or recurrent architectures. Future studies can incorporate AI explainability tools such as Grad-CAM and SHAP in order to enhance transparency and clinical trust. Extending this framework for the multi-class classification and longitudinal tracking of Sarcopenia will enable early detection, progression monitoring, and personalised intervention.

5. Conclusions

This study explored the potential of AI techniques in identifying Sarcopenia through biomechanical analysis of gait videos. By training and comparing several pretrained and custom neural networks, along with the DeepSarcAE model, it was found that deep architectures can effectively distinguish between Sarcopenic and non-Sarcopenic movement patterns using simple two-dimensional video frames. The DeepSarcAE framework is an integration of an autoencoder-based feature extraction module with a supervised CNN classification network to detect Sarcopenia. This achieved the most balanced performance with high sensitivity, specificity, and overall accuracy. The findings suggest that this model can provide a non-invasive, cost-effective, and accessible method for early Sarcopenia screening. This approach requires only a basic video input, making it feasible for the routine monitoring of Sarcopenia without the need for expertise in both clinical and community environments. The study highlights the significance of AI-assisted movement analysis in the diagnosis of ageing populations. Future research directions may focus on developing larger and more diverse datasets, including temporal motion information, and integrating explainable AI methods in order to enhance clinical trust and practical application.

Author Contributions: Conceptualization, M.B. and A.K.D.; Methodology, M.B.; Software, M.B. and A.K.D.; Validation, J.J.M.; Formal analysis, M.B. and A.K.D.; Investigation, J.K. and A.K.D.; Resources, V.K. and A.K.D.; Data curation, M.B. and J.J.M.; Writing—original draft, M.B.; Writing—review and editing, J.J.M. and A.K.D.; Visualisation, J.K.; Supervision, V.K. and A.K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was approved with the ethical clearance number 3075/IEC/2022 from the Human Ethical Committee of the SRM Hospital & Medical College & Research Centre, Kattankulathur, India, on 27 January 2022.

Informed Consent Statement: Written informed consent has been obtained from the patients to publish this paper.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Falcon, L.J.; Harris-Love, M.O. Sarcopenia and the new ICD-10-CM code: Screening, staging, and diagnosis considerations. *Fed. Pract.* **2017**, *34*, 24–32.
2. Rosenberg, I.H. Sarcopenia: Origins and clinical relevance. *J. Nutr.* **1997**, *127*, 990S–991S. [CrossRef]
3. Petermann-Rocha, F.; Gray, S.R.; Pell, J.P.; Balntzi, V.; Lara, J.; Ho, F.K.; Celis-Morales, C. Global prevalence of sarcopenia and severe sarcopenia: A systematic review and meta-analysis. *J. Cachexia Sarcopenia Muscle* **2022**, *13*, 86–99. [CrossRef] [PubMed]
4. Rao, A.R.; Dey, A.B.; Bhagwasia, M.; Singh, V.; Bajpai, S.; Singhal, S.; Chatterjee, P.; Chakrawarty, A.; Dey, S. Prevalence and determinants of sarcopenia among older adults in India: Insights from the Longitudinal Aging Study in India. *J. Am. Geriatr. Soc.* **2025**, *73*, 1429–1440. [CrossRef]
5. Shaikh, N.; Harshitha, R.; Bhargava, M. Prevalence of sarcopenia in an elderly population in rural South India: A cross-sectional study. *F1000Research* **2020**, *9*, 175. [CrossRef]
6. Cruz-Jentoft, A.J.; Bahat, G.; Bauer, J.; Boirie, Y.; Bruyère, O.; Cederholm, T.; Cooper, C.; Landi, F.; Rolland, Y.; Sayer, A.A.; et al. Sarcopenia: Revised European consensus on definition and diagnosis. *Age Ageing* **2019**, *48*, 16–31. Erratum in *Age Ageing* **2019**, *48*, 601. [CrossRef] [PubMed]
7. Chen, L.K.; Woo, J.; Assantachai, P.; Auyeung, T.-W.; Chou, M.-Y.; Iijima, K.; Jang, H.C.; Kang, L.; Kim, M.; Kim, S.; et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. *J. Am. Med. Dir. Assoc.* **2020**, *21*, 300–307.e2. [CrossRef]
8. Buckinx, F.; Landi, F.; Cesari, M.; Fielding, R.A.; Visser, M.; Engelke, K.; Maggi, S.; Dennison, E.; Al-Daghri, N.M.; Allepaerts, S.; et al. Pitfalls in the measurement of muscle mass: A need for a reference standard. *J. Cachexia Sarcopenia Muscle* **2018**, *9*, 269–278. [CrossRef]
9. Takahashi, N.; Sugimoto, M.; Psutka, S.P.; Chen, B.; Moynagh, M.R.; Carter, R.E. Validation study of a new semi-automated software program for CT body composition analysis. *Abdom. Radiol.* **2017**, *42*, 2369–2375. [CrossRef]
10. Supriya, R.; Singh, K.P.; Gao, Y.; Li, F.; Dutheil, F.; Baker, J.S. A multifactorial approach for sarcopenia assessment: A literature review. *Biology* **2021**, *10*, 1354. [CrossRef]
11. Rong, K.; Jiang, L.; Zhou, C.; Yi, X. Intelligent predictive risk assessment and management of sarcopenia in chronic disease patients using machine learning and a web-based tool. *Eur. J. Med. Res.* **2025**, *30*, 345. [CrossRef]
12. Aznar-Gimeno, R.; Alfaro-Santafé, J.; Almenar-Arasanz, A.J.; Perez-Lasierra, J.L.; Pérez-Lázaro, P.; Bosque-López, I.; Azpíroz-Puente, M.; Salvo-Ibáñez, P.; Morita-Hernandez, M.; Hernández-Ruiz, A.C.; et al. Gait-based AI models for detecting sarcopenia and cognitive decline using sensor fusion. *Diagnostics* **2024**, *14*, 2886. [CrossRef]
13. Ko, J.B.; Kim, J.H.; Shin, Y.S.; Han, H.; Han, S.K.; Jung, D.Y.; Hong, J.S. Predicting sarcopenia of female elderly from physical activity performance measurement using machine learning classifiers. *Clin. Interv. Aging* **2021**, *16*, 1723–1733. [CrossRef]
14. Kim, S.H.; Kim, H.S.; Yoo, J. Sarcopenia classification model for musculoskeletal patients using smart insole and artificial-intelligence gait analysis. *J. Cachexia Sarcopenia Muscle* **2023**, *14*, 2793–2803. [CrossRef]
15. Perez-Lasierra, J.L.; Azpíroz-Puente, M.; Alfaro-Santafé, J.V.; Almenar-Arasanz, A.-J.; Alfaro-Santafé, J.; Gómez-Bernal, A. Sarcopenia screening based on the assessment of gait with inertial measurement units: A systematic review. *BMC Geriatr.* **2024**, *24*, 863. [CrossRef]
16. Kim, J.K.; Bae, M.N.; Lee, K.B.; Hong, S.G. Identification of patients with sarcopenia using gait parameters based on inertial sensors. *Sensors* **2021**, *21*, 1786. [CrossRef] [PubMed]
17. Kumar, Y.; Rukhsana. Smart insoles: A cutting-edge technology in the field of orthotics—A review. *Int. J. Health Sci. Res.* **2024**, *14*, 302. [CrossRef]
18. Ruiz-Cárdenas, J.D.; Montemurro, A.; Martinez-García, M.; Rodríguez-Juan, J.J. Sit-to-stand video-analysis app for diagnosing sarcopenia and its relationship with health-related risk factors and frailty in community-dwelling older adults: Diagnostic accuracy study. *J. Med. Internet Res.* **2023**, *25*, e47873. [CrossRef] [PubMed]
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *arXiv* **2015**, arXiv:1512.03385. [CrossRef]
20. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]
21. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A.; Liu, W.; et al. Going deeper with convolutions. *arXiv* **2014**, arXiv:1409.4842. [CrossRef]

22. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]
23. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv* **2020**, arXiv:1905.11946. [CrossRef]
24. Lee, M. GELU activation function in deep learning: A comprehensive mathematical analysis and performance. *arXiv* **2023**, arXiv:2305.12073. [CrossRef]
25. Michelucci, U. An introduction to autoencoders. *arXiv* **2022**, arXiv:2201.03898. [CrossRef]
26. Clevert, D.A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (ELUs). *arXiv* **2016**, arXiv:1511.07289. [CrossRef]