IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 18 September 2023, accepted 7 October 2023, date of publication 16 October 2023, date of current version 20 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3324659

RESEARCH ARTICLE

Ergonomic Risk Prediction for Awkward Postures From 3D Keypoints Using Deep Learning

MD. SHAKHAOUT HOSSAIN[1], SAMI AZAM[2],(Member, IEEE), ASIF KARIM[2], (Member, IEEE), SIDRATUL MONTAHA[3], RYANA QUADIR[4], FRISO DE BOER[2], AND MD. ALTAF-UL-AMIN[1], (Senior Member, IEEE)

[1]Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma 630-0192, Japan
[2]Faculty of Science and Technology, Charles Darwin University, Darwin, NT 0909, Australia
[3]Department of Computer Science, University of Calgary, Calgary, AB T2N 1N4, Canada
[4]Department of Computer Science and Engineering, Daffodil International University, Dhaka 1207, Bangladesh

Corresponding authors: Md. Shakhaout Hossain (hossain.md_shakhaout.hh9@is.naist.jp) and Asif Karim (asif.karim@cdu.edu.au)

ABSTRACT Work-related musculoskeletal ailments are injuries or disorders of the joints, muscles, nerves, or tendons caused by repetitive tasks and jobs that require uncomfortable postures. REBA (Rapid Entire Body Assessment) is a widely used assessment method for examining occupational ergonomics in areas where musculoskeletal disorders (MSDs) are common. REBA assessment necessitates the presence of a professional evaluator who monitors workers' motions and postures, which takes time and has limitations in terms of real-world implementation. With the progress of deep learning-based human posture estimate algorithms, postural risk assessment has become an important and complex research area. We present a technique for forecasting REBA risk levels using 3D coordinates of human body position as input data in this study. We calculated REBA risk scores for various body segments and overall risk rating for corresponding action level for each body position using 3D keypoints from the widely renowned Human 3.6M dataset, which is a significant contribution for future research work in this arena. Using this vast ground truth dataset, a unique DNN model was created to forecast the REBA risk level for measuring the full body's postural risk. REBA Ground Truth dataset is highly imbalanced which coped with data augmentation for the rare classes. To determine the optimal model configuration based on highest accuracy, ablation study is conducted by tuning different hyper-parameters. The proposed model, post-ablation study, attained 89.07% accuracy score on a test set of 128,046 samples from Nadam optimizer with a learning rate of 0.001 and batch size of 512.

INDEX TERMS Ergonomic risk, musculoskeletal disorders (MSDs), 3D-keypoints, posture analysis, rapid entire body assessment (REBA) score.

I. INTRODUCTION

The measurement of workers' postural attitudes in every working environment is critical for assessing and preventing biomechanical overload concerns in the workplace [1]. Musculoskeletal disorders (MSDs) include injuries or abnormalities that affect the muscles, nerves, tendons, joints, cartilage, and spinal discs [2]. Work-related musculoskeletal diseases (WMSD) are highly influenced by the working environment and job performance, and the condition can worsen or last longer as a result of the same work conditions [2]. Any body position that is maintained for an extended amount of time can induce discomfort and exhaustion; for example, standing is a natural body posture that provides no special health risks. However, working for long periods in a standing position can cause discomforts like sore feet, general muscular fatigue, and low back pain [3]. The Bureau of Labor Statistics of the Department of Labor defines MSDs as musculoskeletal system and connective tissue diseases and disorders when the event or exposure

The associate editor coordinating the review of this manuscript and approving it for publication was Giacinto Barresi.

leading to the case is bodily reaction (e.g., bending, climbing, crawling, reaching, twisting), overexertion, or repetitive motion [4]. Employer expenditures related with WMSDs include absenteeism, lost productivity, and increased health care, disability, and worker's compensation costs [5]. The average nonfatal injury or illness is less severe than MSD cases [2]. Around 374 million work-related MSD cases have been documented worldwide, according to the International Labor Organization (ILO). Workers must take an average of four days off work owing to WMSD difficulties. Poor workplace safety and health practises are projected to cost around 3.94% of worldwide gross domestic product per year [6]. Musculoskeletal disorders are common in Australia, with over 6.9 million persons affected by WSMDs in 2014–15, according to the Australian Institute of Health and Welfare [7]. In 2017, 31% of Australians suffer from backpain-related disorders, with 17% of these cases attributed to occupational exposures and risks. According to one study, workplace dangers account for 37% of all back pain worldwide [8]. WMSDs cost more than $24 billion in 2012-2013, according to Safe Work Australia 2015a, and they continue to endanger the work health and safety (WHS) system. These illnesses accounted for 12% of Australia's overall burden of sickness and injury in the general population, and 23% of the non-fatal burden [9]. MSDs are ranked second in terms of their impact based on these numbers. WMSDs have had a significant impact on individuals as well as the economy of the country. Companies must deal with compensation and health-care costs, as well as workers' financial losses as a result of time off or early retirement [10]. Many efficient ergonomic evaluation approaches for analysing the work process have been developed to prevent the incidence of WMSDs. The risk factors for WSMDs and subsequent ergonomic interventions in the workplace are of great consequence, and numerous methodologies such as self-assessment, work posture evaluation and direct measurement [11] have recently been put into action. The process of self-assessment is frequently less reliable and subjective. MSD data and joint angles are typically acquired from workers using different sensors mounted to their bodies [12] while they operate in the direct approach. Some direct techniques necessitate 3D position estimation utilising a depth camera, which has a limited range [13] and is not widely available in workplaces since it requires the configuration of wearable sensors, which disrupts work. The observational evaluation approach has shown to be the most practical method of evaluating postural motions, and it is now widely employed in the industry. It not only allows workers to be viewed immediately and objectively while at work, but it also gives exact and reliable results for MSD risk assessment. Researchers have recently begun to investigate alternative sensor-based automated evaluation approaches, such as distributed surveillance cameras [14], RGB-D cameras [15], reprogrammable Human–Robot Collaboration (HRC) workstations [16], and so on. The most often utilised procedures for measuring ergonomic risks are the Ovako working posture analysis system (OWAS) [17], succinct exposure index (OCRA index) [18], rapid upper limb assessment (RULA) [19], and rapid entire body assessment (REBA) [20]. We focused on postural risk analysis of the full body, not only the upper limb, in this study, thus we used the REBA method, which takes into account all body parts. Deep convolutional networks were the most commonly utilised technique for predicting joint angles and risk assessment (D-CNNs). In this research, we employed a deep learning system to estimate ergonomic risk using 3D human joints keypoints as input.

A. CONTRIBUTIONS

The purpose of this study is to develop a model that reliably forecasts the ergonomic dangers of every difficult posture in any workplace. It is intended that by doing so, the employees' health risks will be reduced, as well as the loss of production. The following key points are worth mentioning:

1) The public dataset used for this study, is a comprehensive one, containing a massive 3.6 million human posture data [21] and ground truth annotation for 17 body joints; we also produced a ground truth dataset with specific body segment scores.
2) Data augmentation was done for training data because the dataset was severely imbalanced, but no augmentation was done for data validation, and all data was taken for testing.
3) The model is sufficiently robust, as validated and evidenced by the necessary ablation studies providing a deep insight into its overall performance.

II. LITERATURE REVIEW

There had been a lot of previous work in the industry using approaches like rapid upper limb assessment (RULA) [19], rapid entire body assessment (REBA) [20], Ovako working posture analysis system (OWAS) [17], postural ergonomic risk assessment (PERA) [22] and succinct exposure index (OCRA index) [18]. During the evaluation process, all of these strategies and their results were assessed, regardless of their various implementation sectors. The RULA method is used to examine the working posture and related risk levels of the upper limb of the body in order to determine MSDs in terms of ergonomic evaluation. Li Li et al. [23] suggested a deep learning-based postural risk factor assessment system based on the RULA methodology for preventing work-related MSD. Normal RGB photos were used to detect 2D poses and generate RULA action levels, which were used to define the RULA grand score category. The work employs data augmentation techniques on a public human pose dataset called Human 3.6. For postures during lifting duties and ordinary activities, their proposed model had a 93% accuracy rate. However, their proposed method could only look at static postures with moderate body motions as a wrist score, and muscle usage was expected to be uniform, although repetitive movement frequency and muscle usage were not taken into account. The suggested algorithm, according to the authors, may not deliver efficient performance in real-world on-site applications because photos often contain noise due to poor lighting conditions, which could lead to overfitting issues. Other posture detectors, RULA estimator structure, and ablation studies on CNN model should also be investigated further to improve the algorithm's resilience. Abobakr et al. [24] later suggested a vision-based semi-automated RULA system for predicting body joint angles from a single depth image. This ergonomic assessment method for adopted working postures uses an inverse kinematics modelling stage to examine the articulated posture by training a deep residual convolutional neural network (CNN). Their proposed method was evaluated in real scenarios with an RULA grand score prediction accuracy of 89%, which was a reasonable agreement; nevertheless, several other evaluation procedures, such as data augmentation and ablation study, might have improved their assessment accuracy. Seo and Lee [25] have suggested a computer vision-based technique to automatically identify construction workers' postures for postural ergonomic risk assessment. By using classification techniques to learn varied postures from virtual images, they were able to minimise the need for large training-image datasets. They used the Support Vector Machine (SVM) as a classifier. This system attained an overall classification accuracy of 89% for automatically analysing ergonomic concerns, however it did not use any established evaluation procedures, and no validation was done. Beheshti et al. [17] proposed using the OVAKO Working posture Analysis System (OWAS) to measure the risk of musculoskeletal problems, and their findings indicated that various farming chores were more risky and hazardous to health. Their findings contrasted risk levels before and after the intervention, as well as the accompanying p-values, but there was no obvious performance measurement or validation. Etemadinezhad et al. [26] used the OWAS approach to investigate musculoskeletal diseases among workers. Their research concentrated on diagnosing damage to various body regions and suggested that there was a pressing need for change. It did not, however, place a strong emphasis on performance measurement or validation of their findings. Wu et. al [11] used REBA technique to obtain a reliable target detection model in 2020 in the form of an App software. This assessment was done based on Mask RCNN (Region Based Convolutional Neural Networks).The effort began with video capture, REBA process implementation, human-machine interaction, video-imaging transmission, and key point extraction. They used the Microsoft COCO dataset and selected 1,000 working photographs from the job site. A total of 800 photos were used as network training data. The developed target detection model was put to the test using 200 test sets. Their model was nearly 90% accurate, but only on a very tiny scale of real-life scenarios. There was no validation process to confirm this assessment, nor was there any ablation study to check the contributions of the components to their model.

Despite the fact that these studies showed a significant increase in terms of predicting postural risk, they were unable to validate the effectiveness of their algorithms. Furthermore, no previous research had conducted an ablation study to determine the important components and the optimal combination for getting the best results.

III. PROPOSED METHODOLOGY

For predicting ergonomic risk with REBA score estimation, this study is proposing a model that stipulates the ergonomic postures that cause musculoskeletal disorders without any human intervention. The dataset used is a public human pose dataset named 'Human 3.6M' for training and evaluation and a ground truth dataset of REBA scores was prepared from the ground-truth keypoints annotation of this public dataset, which we named as REBA DATASET. In this study, after collecting the dataset, the risk level is calculated according to the REBA guide. Afterward, the dataset is distributed based on the five risk levels and split into training, validation and test set. Data augmentation is carried out on the training dataset using three techniques to address the data imbalance issue. To develop the proposed model, we have conducted an ablation study of six cases to get the optimal architecture configuration based on the highest accuracy. The steps of the methodology of the proposed model are shown in Figure 1.

A. DATASET PREPARATION

There are a number of public datasets with ground truth annotations for human body pose. These datasets are annotated for different key-points of human body segments, most of the datasets cover common body segments. Mostly differs for head and neck segments' key-points annotation. For our experiment, we used the Human 3.6M dataset which is one of the largest dataset with both 2D and 3D human pose ground truth key-points for 17 different scenarios. The number of ground truth keypoints data is 2110396 available from this Human3.6m dataset. In Figure 2 one sample annotation for the 17 body keypoints is shown. This dataset is annotated for 17 human body keypoints as follows: **Pelvis, RHip, RKnee, RAnkle, LHip, LKnee, LAnkle, Spine1, Neck, Head, Site, LShoulder, LElbow, LWrist, RShoulder, RElbow, RWrist**

1) GROUND TRUTH (GT) DATASET

We have prepared a ground truth dataset including individual body segments score, score A, score B, score C, Risk and action label for overall posture for both left side and right side of human body posture which can be useful for further research work. The dimension of the GT data is 2110396 $\times$ 75. A snapshot of the ground truth dataset is shown in Table 1 (We have excluded the numerical values of the human body joint coordinates due to the copyright of the Human3.6m dataset).

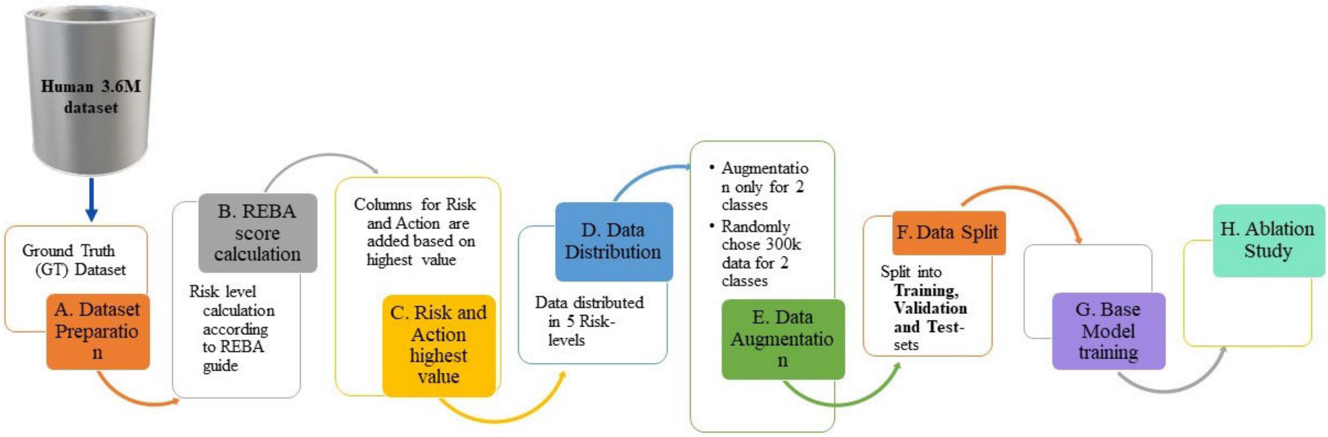

FIGURE 1. Methodology of the proposed model.

TABLE 1. Snapshot of the ground truth dataset.

ImgName	Pelvis_x	Pelvis_y	Pelvis_z	RHip_x	..	Right Action	Risk	Action	Scene
S1/S1_Dir..	-	-	-	-	..	May be necessary	Low	May be necessary	Directions
S1/S1_Dir..	-	-	-	-	..	Necessary	Medium	Necessary	Walk Together

TABLE 2. Ground truth features of Human3.6m dataset.

Features from Human3.6 dataset	ImgName, scene, Pelvis_x, Pelvis_y, Pelvis_z, RHip_x, RHip_y, RHip_z, RKnee_x, RKnee_y, RKnee_z, RAnkle_x, RAnkle_y, RAnkle_z, LHip_x, LHip_y, LHip_z, LKnee_x, LKnee_y, LKnee_z, LAnkle_x, LAnkle_y, LAnkle_z, Spine1_x, Spine1_y, Spine1_z, Neck_x, Neck_y, Neck_z, Head_x, Head_y, Head_z, Site_x, Site_y, Site_z, LShoulder_x, LShoulder_y, LShoulder_z, LElbow_x, LElbow_y, LElbow_z, LWrist_x, LWrist_y, LWrist_z, RShoulder_x, RShoulder_y, RShoulder_z, RElbow_x, RElbow_y, RElbow_z, RWrist_x, RWrist_y, RWrist_z
Features calculated using REBA technique	Left Score A, Left Neck Score, Left Trunk Score, Left Leg Score, Left Score B, Left Upper Arm Score, Left Lower Arm Score, Left Score C, Left Risk, Left Action, RightScore A, Right Neck Score, Right Trunk Score, Right Leg Score, Right Score B, Right Upper Arm Score, Right Lower Arm Score, Right Score C, Right Risk, Right Action

B. REBA SCORE CALCULATION

The data that the GT dataset contains are mostly fetched from the human3.6m dataset and the rest of them are computed using the REBA techniques. The GT features are listed in Table 2. All the left and right scores are calculated using REBA [20] official guidelines. According to the REBA guide, using the Human 3.6M dataset, we can calculate all body segments' risk scores except wrists which we assumed unit value. To calculate the overall risk level for a specific body posture, we first calculate risk for individual body segments and then compute the final REBA score. We have used 3D key-points annotations for REBA score calculation. For this,

1) First, we moved the pelvis key-points to origin (0,0,0) and then transformed the rest of the key-points across the pelvis joint (Figure 3) for the ease of calculation of the segment angles.
2) Next, the angles of the body segments that are required for calculating the REBA score were computed using kinematics from the keypoints (Figure 4).
3) Risk score for each segment was deduced from the REBA [20] guideline using the segment's angle information.
4) Finally, we computed the Grand REBA score using Table C [20] and the corresponding REBA risk label with the help of scores from Table A and Table B [20].
5) Human3.6 dataset contains only various postures not holding any objects. We assumed Load/Force score is zero (< 5 kg exerted). So we added Load/Force score zero to get the final score of Table A.
6) We added Coupling score (0) with the Table B output score. Coupling is assumed Good (0) which is Well-fitting handle and a mid-range, power grip, as these information can not be extracted from the dataset.
7) Activity score can be derived by analysing continuous video of a person. As we have used single frame for computing REBA score, we have ignored the Activity score from the calculation.

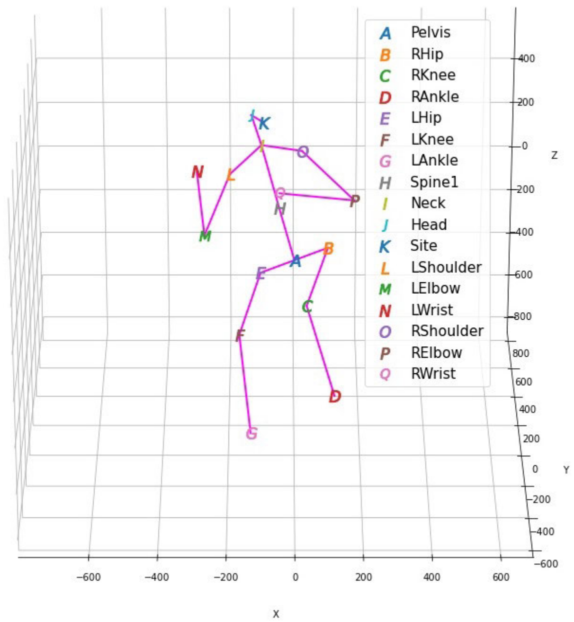

FIGURE 2. The 17 human body keypoints annotation.

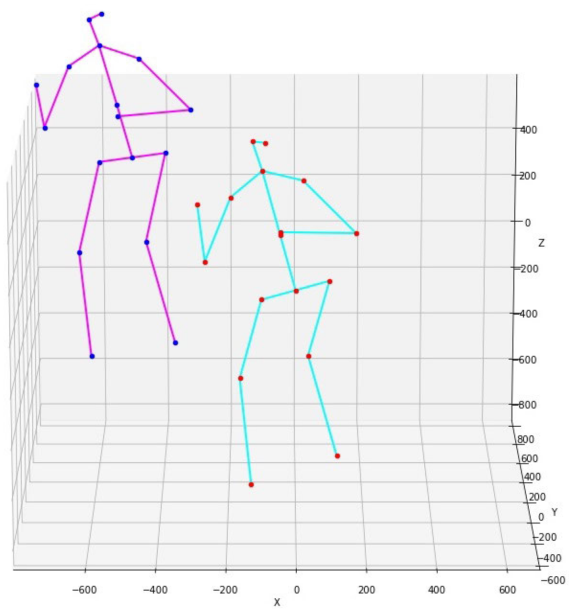

FIGURE 3. Ground Truth pose moved across Origin.

In this work, we only used 51 coordinates of 17 body joint key-points and corresponding risk levels for each posture.

C. RISK AND ACTION CALCULATION

The features- Risk and Action columns are the highest value from corresponding 'Left Risk' and 'Right Risk' and 'Left Action' and 'Right Action' features, as we are only concerned about the highest score for any posture. This has a huge impact in the model since sometimes some models predict the lower class value. In the ground truth dataset, we have 1079946 samples in the train set, 186542 samples in the validation set and 186484 samples in the test set.

TABLE 3. Data distribution in different scenes.

Scene Type	Number of Samples
Discussion	223032
Directions	134664
Eating	148732
Greeting	103080
Phoning	171920
Photo	96928
Posing	96748
Purchases	82364
Sitting	156748
SittingDown	162556
Smoking	188804
TakingPhoto	8440
Waiting	153160
WalkDog	93960
WalkTogether	113496
Walking	162008
WalkingDog	13756

D. DATA DISTRIBUTION

According to REBA guidelines [20], there are five risk levels: Negligible, Low, Medium, High and Very High. Ground truth dataset is prepared from Human 3.6M 3D keypoints by calculating segments angle and REBA tables. As Human 3.6M contains daily postures, the ground truth REBA risk levels derived from it is highly skewed. For 'Medium' risk labels there are almost more than 60% of data keypoints, where 'Low' and 'High' risk class have sufficient data about 250231 and 452564. But for 'Negligible' and 'Very High' risk classes there are only 475 and 2929 data keypoints which are even less than 1% of the full dataset. Figure 5 explains the data distribution of our proposed dataset in terms of Risk-level.

Along with keypoints information in Human 3.6M dataset there are annotations for scene types. The dataset has some specific Scene-types (e.g., eating, greeting, sitting, taking photos etc.) and the data distribution Human 3.6M for different scene types are shown in Table 3.

E. DATA SPLIT

As the REBA ground truth dataset is highly imbalanced as shown in figure 5, randomly chose 300000 samples for 'Medium' & 'High' classes. A total of 853635 samples out of 2110396 were taken for model training, validation and testing purposes. Then the dataset is split as 70% training, 15% validation and 15% test data.

F. DATA AUGMENTATION

From figure 5, it can be noticed that in two specific classes- 'Negligible' and 'Very High', there was the least number of samples and these classes are quite rare. As this is a classification problem with this highly skewed data, majority classes will suffer from overfitting while underfitting for minority classes. To cope with this problem, we have adopted data augmentation techniques for the two minority

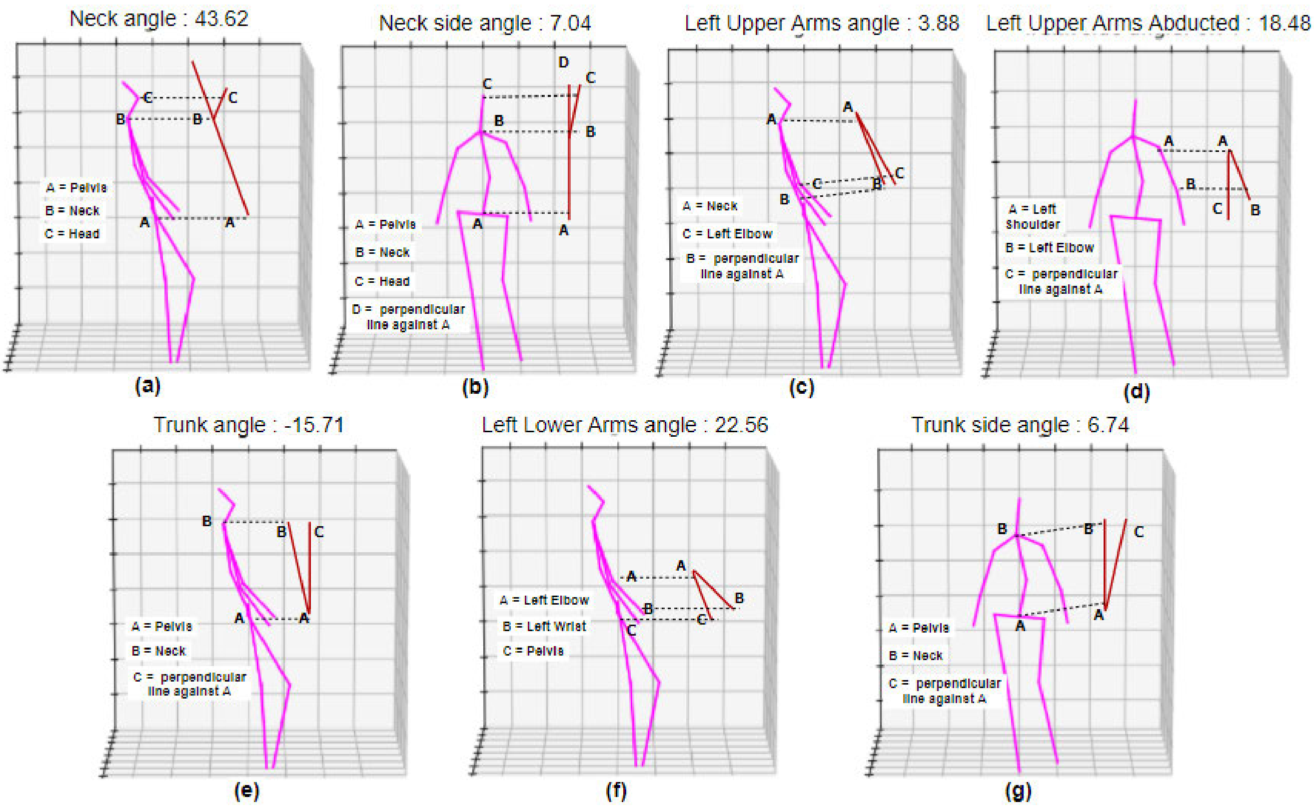

FIGURE 4. Segments' angles were calculated using Kinematics from the 3D keypoints. Each segment angle was used to calculate corresponding REBA risk score.

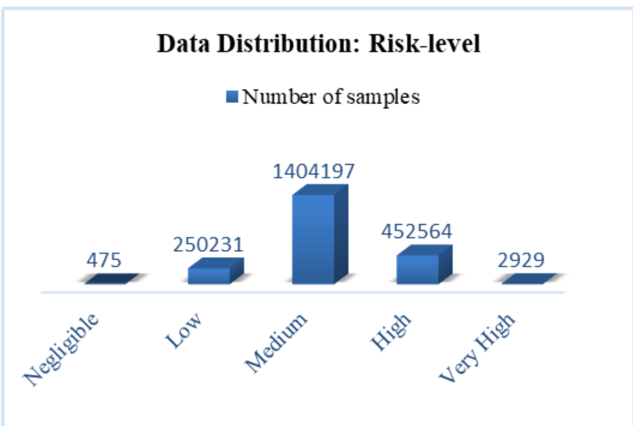

FIGURE 5. Data distribution in different risk-levels.

classes only for model training. Three types of augmentation techniques were used, namely: Pose rotation, keypoint translation and multiplying the pose with a certain multiplier. In the keypoint translation technique, each joint coordinate is translated from its previous state. In Figure 6(ii) only the left wrist keypoint translation technique is shown. In the third data augmentation technique we have multiplied all keypoints with a certain multiplier. In Figure 6(iii) augmented pose shrunk due to multiplying with a certain multiplier.

TABLE 4. Train dataset after data augmentation.

Data Type	Negligible	Low	Medium	High	Very High
Train data(after aug-menta-tion)	18860	175162	210000	210000	32800

In the pose rotation technique, the whole human pose is rotated with respect to the right hip joint as shown in Figure 6(iv). After applying all three types of augmentation techniques, computed REBA score for the augmented data. For some augmented data, REBA scores changed from their original scores. Removed the data which changed from the previous labels after augmentation. For model validation and testing, no data augmentation was performed. After the Data Augmentation, the training dataset appears as Table 3.

G. BASE MODEL

The purpose of this study is to predict REBA risk levels from 3D joint coordinates for human postures. For estimating the REBA risk level, we made a DNN (Deep Neural Network) model. The input to the DNN model is a 51×1 vector of the human 17 joints key-points, and the output is a 5×1 vector

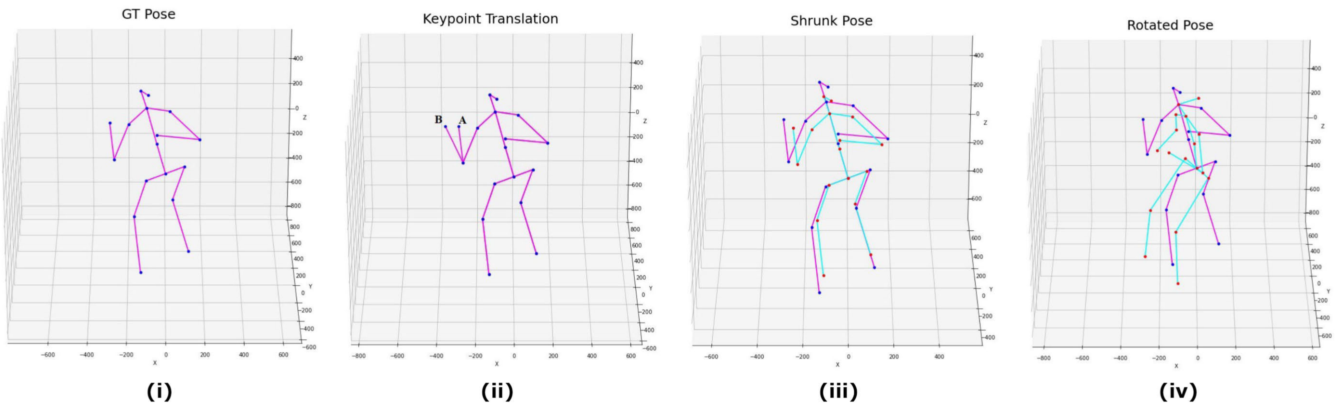

FIGURE 6. Data augmentation techniques i) Ground truth, ii) Keypoint translation, iii) Shrunk pose, iv) Pose rotation.

TABLE 5. Reba action levels.

Action level	REBA score	Risk level	Action (including further assessment)
0	1	Negligible	None necessary
1	2-3	Low	May be necessary
2	4-7	Medium	Necessary
3	8-10	High	Necessary soon
4	11-15	Very High	Necessary now

corresponding to the five risk levels based on the REBA grand score (Table 4).

Initially we trained a base model for 100 epochs. In the base model, 7 dense layers used with neurons (256, 256, 128, 128, 64, 64, 32) and LeakyRelu as activation function with alpha value 0.50 were used. In the hyperparameters batch size of 512, learning rate 0.001 and Adam optimizer were used. The base model yielded a 0.7744 accuracy score in the test set.

H. ABLATION STUDY

Ablation study is carried out to achieve the optimal configuration of the model based on the highest accuracy by altering several components and hyperparameters [27], [28]. Results and findings of the complete ablation study having seven study cases are showcased in Table 5 and Table 6. Table 5 covers all the results regarding the model's layer architecture of different configurations and activation functions whereas Table 6 comprises the results of tweaking several hyperparameters, loss function and flatten layer. Over the seven study cases, a significant improvement in test accuracy from 80.13% to 89.39% is achieved gradually.

1) CASE STUDY 1: ALTERING THE NUMBER OF HIDDEN LAYER

In this case study, the experimentation is carried out by changing the number of hidden layers. To begin with, the number of hidden layers is set to 5 which are increased afterwards gradually to perceive the performance. It is observed from Table 5 that the best performance is attained from configuration 3 having 7 hidden layers with the highest test accuracy of 81.30% in 188 epochs. Though configuration 1 took the lowest number of epochs to yield the best performance, the test accuracy is not equitable enough compared to configuration 3. Therefore, this configuration of 7 hidden layers is picked to proceed with the forthcoming case studies.

TABLE 6. Ablation study regarding layer configurations and activation functions.

Case study 1: Changing Hidden Layer				
Conf. No.	No. of Hidden layer	Epoch	Test accuracy (%)	Finding
1	5	178	80.13	Lowest accuracy
2	6	216	80.69	Modest accuracy
3	**7**	**188**	**81.30**	**Highest accuracy**

Case study 2: Changing the number of neurons				
Conf. No.	No. of Neuron	Epoch	Test accuracy (%)	Finding
1	256à 256à 128à 128à 64à 64à 32	188	81.30	Previous accuracy
2	256à 128à 128à 128à 64à 64à 64	230	80.78	Accuracy dropped
3	128à 128à 128à 128à 64à 64à 64	260	80.90	Accuracy dropped
4	128à 128à 128à 128à 64à 64à 32	261	81.37	Accuracy improved
5	**256à 128à 128à 128à 64à 64à 32**	**219**	**81.61**	**Highest accuracy**

Case study 3: Changing activation function				
Conf. No.	Activation function	Epoch	Test accuracy (%)	Finding
1	**Leaky ReLu**	**219**	**81.61**	**Previous accuracy**
2	Relu	169	77.96	Accuracy dropped
3	PReLU	229	80.03	Accuracy dropped

TABLE 7. Ablation study regaring model hyper-parameters, loss function and flatten layer.

Case study 4: Changing Optimizer				
Conf. No.	Optimizer	Epoch	Test accuracy (%)	Finding
1	SGD	219	81.61	Previous accuracy
2	Adam	371	88.02	Accuracy improved
3	**Nadam**	**372**	**88.77**	**Highest accuracy**

Case Study 5: Changing Batch Size				
Conf. No.	Batch Size	Epoch	Test accuracy (%)	Finding
1	256	372	88.77	Previous accuracy
2	**512**	**378**	**89.07**	**Highest accuracy**
3	1024	314	89.02	Accuracy improved

Case study 6: Changing Learning Rate				
Conf. No.	Learning Rate	Epoch	Test accuracy (%)	Finding
1	**0.001**	**378**	**89.07**	**Previous accuracy**
2	0.01	288	87.64	Accuracy dropped
3	0.0005	302	88.68	Accuracy dropped

2) CASE STUDY 2: ALTERING THE NUMBER OF NEURONS

The number of neurons in each hidden layer yields prominent consequences on a model's performance. Hence, the proposed model has been experimented five times with five configurations of the number of neurons. For 7 hidden layers, the configurations are generated by decreasing the number of neurons on each layer gradually. It is observed that for configuration 5 (256Ã 128Ã 128Ã 128Ã 64Ã 64Ã 32) which comprises 256, 128, 128, 128, 64, 64 and 32 neurons for the 7 hidden layers records the utmost performance with a test accuracy of 81.61%. Moreover, the total number of 219 epochs was required to reach the performance which is the second lowest among all configurations. Therefore, we move ahead with further case studies using configuration 5.

3) CASE STUDY 3: ALTERING ACTIVATION FUNCTION

Since different activation functions have an effect on the overall performance, selecting an optimum activation function based on the best performance is a significant research consideration. In this case study, experimentation of three activation functions named PReLU, ReLU and Leaky ReLU, are carried out where Leaky ReLu outperforms others acquiring the identical test accuracy of 81.61%. This activation function is chosen for further ablation studies.

TABLE 8. Configuration of the optimal model architecture.

Configuration	Value
No. of Hidden layer	7
No. of Neuron	256à 128à 128à 128à 64à 64à 32
Activation function	Leaky ReLu
Optimizer	Nadam
Batch size	512
Learning rate	0.001
Epoch	378

4) CASE STUDY 4: CHANGING OPTIMIZERS

Testing a model with various optimizers can positively impact the performance and maximize testing accuracy. In this regard, a total of three optimizers namely SGD, ADAM and Nadam are experimented with to better configure the proposed model. It is evident from Table 6 that changing the optimizer from SGD to Adam and Nadam has significantly increased the testing accuracy from 81.61% to above 88%. However, the best performance is achieved from Nadam optimizer with a test accuracy of 88.77% hence this optimizer is chosen for further ablation studies.

5) CASE STUDY 5: CHANGING BATCH SIZE

The number of batch sizes during training of a deep learning model can slightly impact the overall performance of a model. In order to achieve the best configuration of the proposed model, experimentations with batch sizes 256, 512 and 1024 are conducted. In this regard, the best performance is achieved from batch size 512 with a test accuracy of 89.07%. Thus, this batch size is chosen for further case studies.

6) CASE STUDY 6: CHANGING LEARNING RATE

Different learning rates can impact the training phases of a model and have an influence on the overall performance of the model. In order to configure the model with the most optimized learning rate, experimentation with a total of three learning rates is conducted. Among them, the previously employed learning rate of 0.001 outperformed the other learning rates by maintaining a test accuracy of 89.07%. As a result, this learning rate is chosen for the proposed model.

After the conduction of ablation study consisting of six test cases, the configuration of the best resultant architecture is shown in Table 7.

I. MSD ESTIMATOR

After ablation study, the MSD Estimator model is chosen. The model architecture of the MSD Estimator is shown in figure 7. Model has 7 Dense layers with neurons (256, 128, 128, 128, 64, 64, 32), LeakyRelu with alpha value 0.50 and initial number of epochs is set to 500. Other hyperparameters are Nadam optimizer, batch size 512 and learning rate 0.001. In the model, regularization techniques are used to check overfitting like Batch Normalization after applying each Leaky ReLU and also Dropout with probability 20% used after the Batch Normalization of the first Dense

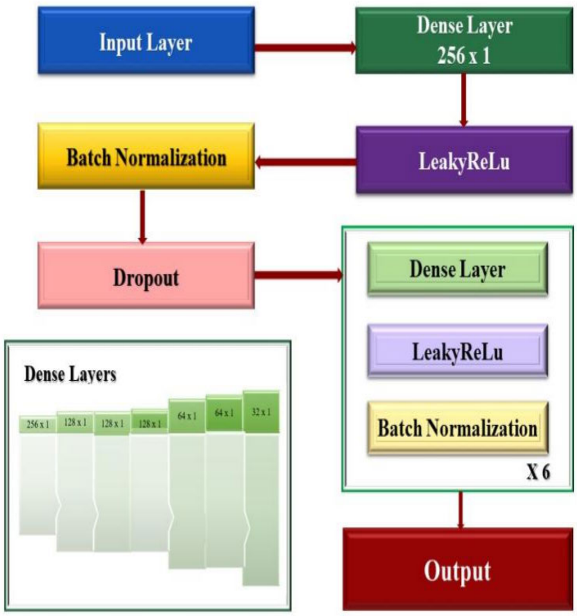

FIGURE 7. Model Architecture of the MSD Estimator.

TABLE 9. Classification report.

Risk Classes	Precision	Recall	F1-Score	Support
Negligible	0.40	0.99	0.57	71
Low	0.89	0.94	0.91	37535
Medium	0.89	0.80	0.854	45000
High	0.91	0.94	0.92	45000
Very High	0.57	0.98	0.72	440
Accuracy			0.8907	128046
Macro Avg.	0.73	0.93	0.79	128046
Weighted Avg.	0.89	0.89	0.89	128046

TABLE 10. Confusion matrix.

True \ False	Negligible	Low	Medium	High	Very High
Negligible	70	1	0	0	0
Low	93	35169	2261	12	0
Medium	9	4489	36123	4376	3
High	1	5	2411	42257	326
Very High	0	0	0	10	430

layer. We used early stopping with monitoring validation accuracy and stop training if accuracy did not improve within 25 epochs for checking overfitting of the model. We also used ReduceLROnPlateau to reduce the learning rate with a factor 0.7, patience 10 and minimum learning rate 0.000001 by monitoring validation accuracy. Though the model is trained for 378 epochs, saved the best weight for the maximum validation accuracy using the model checkpoint parameter.

IV. RESULT

REBA ground truth dataset was prepared using a kinematics formula and verified the REBA score computation by manually checking and visualizing the scores for each body segment taking large samples from each class. According to Table 3 data distribution some scene types samples are not relevant for this dataset. Samples from Eating, Sitting, Taking Photo, Waiting etc have some unusal postures which were removed by checking some segments angle. Some postures like lying can be easily determined by the trunk angle or lower legs angle with respect to the ground were removed. Also we took each body segments angle distribution and removed the samples those fall beyond our assumed thresholds. Our model achieved 89.07% overall accuracy over a test set of 128046 samples which is quite promising. In Table 8, the classification report on the test dataset is shown. For 'Medium' class recall rate is low which is 80% whereas for other classes at least 94% or more.

Though 'Negligible' and 'Very High' classes have a low number of samples, the highest recall rate was achieved due to data augmentation. Confusion matrix is shown in Table 10 where we can see that most of the miss classified classes are in neighboring classes. The reason behind this is, for the slight changes in ergonomic postural risk level changes from one risk class to another. Some of the detected class is not even near the actual class. Human 3.6 million dataset has various types of body postures. Some of the postures are not even related to this task, suppose we do not want to measure the MSD risk while a person is lying on the floor. For these types of unusual body postures MSD risk score will be different as there need some assumptions. In Human 3.6m dataset, there are a number of unusual data which can be one reason for the far class prediction. As this dataset is very large we could not remove these outlier data.

V. DISCUSSION

In this study, we proposed a method to predict the ergonomic risk of a posture. We used rapid entire body assessment (REBA) to assess the ergonomic postures. Manually assessing REBA risk level needs an expert evaluator and is also not feasible for large product lines. Our proposed method uses 3D human joints keypoints for the REBA risk assessment system. There are some available algorithms to extract 3D human keypoints form RGB images. These 3D human coordinates are used as the input to our DNN model which predicts the risk level for the posture. In multiclass classification for the 5 risk classes, our model achieved 89.07% overall accuracy and 89% weighted average f1-score over a large test set from the Human 3.6m public dataset. For the 'Medium' class, this model got relatively low accuracy in both training and validation dataset and most of the misclassified data are in 'Low' and 'High' classes. As a slight change in angle of the body segments shifts the risk from one level to another and in actual data 'Medium' class had more than 66% data samples, the model yielded the lowest accuracy for this class. According to the augmentation technique used for the rare classes and their accuracy score, it is noticeable that data

TABLE 11. Comparison of the prior and proposed work.

Author	Dataset	Assessment Method	Algorithm	Performance
Ahmed Abobakr, Darius Nahavandi et al. [24]	ImageNet	RULA	CNN	MAE = 4.4%
Li Li, Tara Martin, Xu Xu et al. [23]	Custom dataset	RULA	CNN	Accuracy = 93%
JoonOh Seo and SangHyun Lee et al. [25]	Custom dataset	OWAS	SVM	Accuracy = 89%
Shuai Wu, Ziwen Chen et al. [11]	Custom dataset	REBA	Mask RCNN	–
Chatzis T, Konstantinidis D, Dimitropoulos K et al. [29]	UW-IOM	REBA	VAE	MAE = 0.377 ± 0.04
Proposed System	**Human 3.6m**	**REBA**	**DNN**	**Accuracy = 89.07%**

augmentation for the 'Medium' class would produce a good accuracy score. This is a large scale dataset and the data is not balanced at all. Having only few samples for the rare classes and achieving this much accuracy was really challenging, where most of the works showed their achievements on limited scale custom dataset.

From the prior work described in the literature review, it is clearly visible that using the REBA guide, no other work showed such high performance using this massive size dataset in their work. Table 11 can clearly illustrate this comparison. Of the two REBA based techniques, first approach used Mask RCNN algorithm to extract the keypoints and then using kinematics computed the REBA score using an application. In their study they compared their application's REBA score with human determined scores on a custom dataset. The other technique used UW-IOM dataset in their study and using multiple Variational Aligning Process predicted REBA score with MAE 0.377 ± 0.04. In UW-IOM dataset there are not even any data sample for Negligible class. In our study we have used the largest dataset with all the REBA classes and achieved competitive accuracy score using very simple DNN architecture compared to the other studies where they used very complex CNN architectures. Another advantage of using this simple DNN architecture is that due to faster computation it is suitable for real time implementation. This work is a step-stone for future research work. As this dataset covers all the risk classes, this work can be extended further for practical implementation. In factories or work places where labor intensive work is being conducted, this automatic risk predictor can be implemented by getting data from cameras present in the places which can tackle MSDs of the workers by correcting their work postures.

TABLE 12. Posture Scores used in the determination of REBA Score.

Table A													
Trunk		Neck											
		1				2				3			
1	Legs	1	2	3	4	1	2	3	4	1	2	3	4
2		1	2	3	4	1	2	3	4	3	3	5	6
3		2	3	4	5	3	4	5	6	4	5	6	7
4		3	5	6	7	5	6	7	8	6	7	8	9
5		4	6	7	8	6	7	8	9	7	8	9	9

Table B			
		Lower Arm	
Upper Arm		1	2
	Wrist	1	1
1		1	1
2		1	2
3		3	4
4		4	5
5		6	7
6		7	8

	Table C								
	Score B								
Score A		1	2	3	4	5	6	7	8
	1	1	1	1	2	3	3	4	5
	2	1	2	2	3	4	4	5	6
	3	2	3	3	3	4	5	6	7
	4	3	4	4	4	5	6	7	8
	5	4	4	4	5	6	7	8	8
	6	6	6	6	7	8	8	9	9
	7	7	7	7	8	9	9	9	10
	8	8	8	8	9	10	10	10	10
	9	9	9	9	10	10	10	11	11

VI. CONCLUSION AND FUTURE WORKS

MSDs in the workplace can cause nonfatal injuries and illnesses such as Tendonitis, Epicondylitis, Carpal tunnel syndrome, DeQuervain's disease, Thoracic outlet syndrome, Tension neck syndrome, and others. Millions of people around the country take time off work to manage and recuperate from work-related musculoskeletal discomfort or disability in the lower back or upper extremities. This methodology, which is based on the REBA method, can be used for any workplace postural risk assessment, regardless of the work situation. We used one of the largest datasets of human position, which we then produced with body segment scores for our purposes. We estimated risk for individual body segments first, and then computed the total REBA score to determine the overall risk level for a certain body posture. The REBA score was calculated using 3D key-point annotations.

We have compiled a substantial ground truth dataset that can be used in a variety of ergonomic posture analysis studies. This dataset comprises of all the risk classes that can be extended to implement in real life scenarios where complex postures are involved. The ablation study assisted us in achieving the best outcome by adjusting the model's components to achieve the best result. In such a huge dataset, the accuracy and weighted average were both satisfactory at 89.07 percent. Using comparable data augmentation for the other classes could improve accuracy, which could be the subject of future research. We plan to develop a vision-based system that can detect 3D human body keypoints from a single RGB image in the future. Our current MSD Estimator algorithm will use these extracted keypoints as input to predict postural risk as a REBA risk level.

APPENDIX

See Table 12.

ACKNOWLEDGMENT

The authors would like to thank the contributions of the authors of Human3.6m dataset for developing and maintaining such a well organized dataset.

REFERENCES

[1] Z. Li, R. Zhang, C.-H. Lee, and Y.-C. Lee, "An evaluation of posture recognition based on intelligent rapid entire body assessment system for determining musculoskeletal disorders," *Sensors*, vol. 20, no. 16, p. 4414, Aug. 2020.

[2] *Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities*, National Research Council and Institute of Medicine, USA, 2001.

[3] Canadian Centre for Occupational Health & Safety (CCOHS). (2019). *Work-Related Musculoskeletal Disorders (WMSDS)—Risk Factors*. [Online]. Available: https://www.ccohs.ca/oshanswers/ergonomics/risk.html

[4] M. W. Brault, J. Hootman, C. G. Helmick, K. A. Theis, and B. S. Armour, "Prevalence and most common causes of disability among adults–United States, 2005," *Morbidity Mortality Weekly Rep.*, vol. 58, no. 16, pp. 421–426, 2009.

[5] *Work-Related Musculoskeletal Disorders & Ergonomics*, Centers for Disease Control and Prevention (CDC), USA, 2020.

[6] International Labor Organization. *Safety and Health at Work*. Accessed: May 5, 2023. [Online]. Available: https://www.ilo.org/global/standards/subjects-covered-by-international-labour-standards/occupational-safety-and-health/lang–en/index.htm

[7] J. Oakman, S. Clune, and R. Stuckey, "Work-related musculoskeletal disorders in Australia," Safe Work Australia, Canberra, ACT, Australia, Tech. Rep., 2019. [Online]. Available: https://www.safeworkaustralia.gov.au/system/files/documents/1912/work-related_musculoskeletal_disorders_in_australia_0.pdf

[8] M. Fingerhut, T. Driscoll, M. Concha-Barrientos, L. Punnett, A. Pruss-Ustin, K. Steenland, J. Leigh, and C. Corvalan, "Contribution of occupational risk factors to the global burden of disease—A Summary of findings," *SJWEH Supplements*, vol. 2005, no. 1, pp. 58–61, 2005.

[9] *Musculoskeletal Disorders: Comprehensive Approach Needed*, Safety Solutions, MMWR, USA, 2020.

[10] R. Jonsson, L. Dellve, and B. Halleröd, "Work despite poor health? A 14-year follow-up of how individual work accommodations are extending the time to retirement for workers with poor health conditions," *SSM-Population Health*, vol. 9, Dec. 2019, Art. no. 100514.

[11] S. Wu, Z. Chen, X. Zhao, M. Yao, Z. Wang, and S. Kuang, "Design of an ergonomic app for entire rapid body assessment based on mask RCNN," *J. Phys., Conf. Ser.*, vol. 1633, no. 1, Sep. 2020, Art. no. 012150.

[12] H.-C. Chen, Y.-P. Liu, W.-H. Hong, Y.-C. Lin, and C.-Y. Yu, "Evaluation of a proposed chair with an arm support for wiring terminal blocks on a vertical plane," *Ind. Health*, vol. 57, no. 3, pp. 359–369, 2019.

[13] A. Seuser, K. Kurnik, and A.-K. Mahlein, "Infrared thermography as a non-invasive tool to explore differences in the musculoskeletal system of children with hemophilia compared to an age-matched healthy group," *Sensors*, vol. 18, no. 2, p. 518, Feb. 2018.

[14] C. Li and S. Lee, "Computer vision techniques for worker motion analysis to reduce musculoskeletal disorders in construction," in *Computing in Civil Engineering*. USA: ASCE Press, 2011, pp. 380–387.

[15] A. Shafti, A. Ataka, B. U. Lazpita, A. Shiva, H. A. Wurdemann, and K. Althoefer, "Real-time robot-assisted ergonomics," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 1975–1981.

[16] W. Kim, M. Lorenzini, P. Balatti, P. D. H. Nguyen, U. Pattacini, V. Tikhanoff, L. Peternel, C. Fantacci, L. Natale, G. Metta, and A. Ajoudani, "Adaptable workstations for human–robot collaboration: A reconfigurable framework for improving worker ergonomics and productivity," *IEEE Robot. Autom. Mag.*, vol. 26, no. 3, pp. 14–26, Sep. 2019.

[17] M. H. Beheshti, A. FirooziChahak, A. A. AlinaghiLangari, and M. Poursadeghiyan, "Risk assessment of musculoskeletal disorders by OVAKO working posture analysis system OWAS and evaluate the effect of ergonomic training on posture of farmers," *J. Occupational Health Epidemiol.*, vol. 4, no. 3, pp. 131–138, 2015.

[18] E. Occhipinti and D. Colombini, "Updating reference values and predictive models of the OCRA method in the risk assessment of work-related musculoskeletal disorders of the upper limbs," *Ergonomics*, vol. 50, no. 11, pp. 1727–1739, Nov. 2007.

[19] L. McAtamney and N. Corlett, "Rapid upper limb assessment (RULA)," in *Handbook of Human Factors and Ergonomics Methods*. U.K.: Routledge, 2004.

[20] S. Hignett and L. McAtamney, "Rapid entire body assessment (REBA)," *Appl. Ergonom.*, vol. 31, no. 2, pp. 201–205, Apr. 2000.

[21] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014.

[22] D. S. Chander and M. P. Cavatorta, "An observational method for postural ergonomic risk assessment (PERA)," *Int. J. Ind. Ergonom.*, vol. 57, pp. 32–41, Jan. 2017.

[23] L. Li, T. Martin, and X. Xu, "A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders," *Appl. Ergonom.*, vol. 87, Sep. 2020, Art. no. 103138.

[24] A. Abobakr, D. Nahavandi, J. Iskander, M. Hossny, S. Nahavandi, and M. Smets, "A Kinect-based workplace postural analysis system using deep residual networks," in *Proc. IEEE Int. Syst. Eng. Symp. (ISSE)*, Oct. 2017, pp. 1–6.

[25] J. Seo and S. Lee, "Automated postural ergonomic risk assessment using vision-based posture classification," *Autom. Construct.*, vol. 128, Aug. 2021, Art. no. 103725.

[26] S. Etemadinezhad, F. Ranjbar, and M. Gorji, "Posture analysis by OWAS method and prevalence of musculoskeletal disorders using Nordic questionnaire among workers of Sourak tobacco factory in 2013," *Iranian J. Health Sci.*, vol. 1, no. 2, pp. 89–94, Sep. 2013.

[27] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, "Ablation studies in artificial neural networks," 2019, *arXiv:1901.08644*.

[28] S. Montaha, S. Azam, A. K. M. R. H. Rafid, Md. Z. Hasan, A. Karim, and A. Islam, "TimeDistributed-CNN-LSTM: A hybrid approach combining CNN and LSTM to classify brain tumor on 3D MRI scans performing ablation study," *IEEE Access*, vol. 10, pp. 60039–60059, 2022.

[29] T. Chatzis, D. Konstantinidis, and K. Dimitropoulos, "Automatic ergonomic risk assessment using a variational deep network architecture," *Sensors*, vol. 22, no. 16, p. 6051, Aug. 2022.

MD. SHAKHAOUT HOSSAIN received the B.Sc. degree in mechanical engineering from the Bangladesh University of Engineering and Technology, Bangladesh. He is currently pursuing the Ph.D. degree in information science with the Nara Institute of Science and Technology (NAIST), Japan. Along with his Ph.D. research, he is a Lead AI Engineer with Hiperdyne Corporation, Tokyo, Japan. His research interests include computer vision, machine learning in medical imaging, and MLOps.

SAMI AZAM (Member, IEEE) is currently a Leading Researcher and a Senior Lecturer with the College of Engineering and IT, Charles Darwin University, Casuarina, NT, Australia. He has several publications in peer-reviewed journals and international conference proceedings. His research interests include computer vision, signal processing, artificial intelligence, and biomedical engineering.

ASIF KARIM (Member, IEEE) is currently a Research Active Lecturer with Charles Darwin University, Australia. Alongside being an Active Researcher, he has considerable industry experience in IT, primarily in the field of software engineering. His research interests include machine intelligence, health informatics, and smart contracts.

SIDRATUL MONTAHA received the B.Sc. degree in computer science and engineering from Daffodil International University, Dhaka, Bangladesh. She is currently pursuing the master's degree with the University of Calgary. Her research interests include computer vision, health informatics, machine learning, deep learning, and artificial intelligence-based systems.

RYANA QUADIR received the B.Sc. degree in computer science and engineering from North South University (NSU), and the master's degree from Daffodil International University (DIU), Bangladesh. She was a Software QA Automation Engineer with Stibo DX, for seven years. Her research interests include machine learning and fuzzy systems and AI in medical diagnostics.

FRISO DE BOER is currently a Professor with the Faculty of Science and Technology, Charles Darwin University, Casuarina, NT, Australia. His research interests include signal processing, biomedical engineering, and mechatronics.

MD. ALTAF-UL-AMIN (Senior Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology (BUET), Dhaka, the M.Sc. degree in electrical, electronic and systems engineering from Universiti Kebangsaan Malaysia (UKM), and the Ph.D. degree from the Nara Institute of Science and Technology (NAIST), Japan. He is currently an Associate Professor with the Computational Systems Biology Laboratory, NAIST. His research interests include network biology, systems biology, cheminformatics, and biological databases.