International Journal of Intelligent Information Systems
2022; 11(1): 1-6
http://www.sciencepublishinggroup.com/j/ijiis
doi: 10.11648/j.ijiis.20221101.11
ISSN: 2328-7675 (Print); ISSN: 2328-7683 (Online)

A Multiple Classification Model for the Prediction of Work-Related Musculoskeletal Disorder

Amadi Chimeremma Sandra[1], John-Otumu Adetokunbo Macgregor[1], Eze Peter Uchenna[2]

[1]Department of Information Technology, Federal University of Technology, Owerri, Nigeria

[2]Department of Computiing and Information Systems, University of Melbourne, Victoria, Australia

Email address:

amadi.sandra@futo.edu.ng (Amadi C. S.), adetokunbo.johnotumu@futo.edu.ng (John-Otumu A. M.)

To cite this article:

Amadi Chimeremma Sandra, John-Otumu Adetokunbo Macgregor, Eze Peter Uchenna. A Multiple Classification Model for the Prediction of Work-Related Musculoskeletal Disorder. *International Journal of Intelligent Information Systems*. Vol. 11, No. 1, 2022, pp. 1-6.
doi: 10.11648/j.ijiis.20221101.11

Received: December 3, 2021; **Accepted**: December 23, 2021; **Published**: January 8, 2022

Abstract: In recent times, musculoskeletal disorders (MSD) represent one of the most common and expensive occupational health problems in both developed and developing countries. Work-related musculoskeletal disorders (WRMSD) are impairments that are mostly caused by the workplace and immediate environment. A two-step predictive model is introduced here using KNN and Decision tree machine learning algorithms. This model for predicting WRMSD enables for early detection and correction of upper and lower back disorders, carpal tunnel syndrome and other WRMSD disorders associated with office workers. Key informant interview technique, observation of previous methods, online repository and published related works were used in data gathering. In training the model, 80% of the dataset was used while 20% was used for testing the model to prevent overfitting using python programming language. JavaScript, Hypertext preprocessor (PHP), Hypertext Markup Language (HTML), Cascading Stylesheet (CSS) and MySQL were also used to develop the front and backend of the application. The result revealed that the proposed model had 90.44% accuracy, 92.71% Recall (sensitivity), 97.16% precision, and 94.88% F1-Score. The proposed model, however, makes it easy for multiple classifications in other to predict both present and future risk of WRMSD. Performance is estimated to have high accuracy, recall, precision and f1 score in comparison to other existing algorithms.

Keywords: WRMSD, Decision Tree, K-NN, Classification, Machine Learning, Predictive Model

1. Introduction

The practice and application of office ergonomics is not a new phenomenon in today's world. The work environment is assumed to have a significant influence on better work outcomes and productivity [1].

Ergonomics can be defined as the scientific study of the link between a person and his or her working conditions. According to the International Ergonomics Association (IEA), ergonomics or human factor is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and methods to design in other to optimize human well-being and overall system performance [2, 3]. The principles of ergonomics are broadly extended to address other user environments such as seen in healthcare systems [4], agriculture [5], and recreational industries [6], However, the problem of ergonomic hazard among office workers is very extensive and broad, some studies have shown significant prevalence rates among those in the Nigerian civil service with MSDs mostly in the lower back region [7]. As with any problem, there is an intervention; preventing ergonomic hazard risk factors can be accomplished in a variety of ways, including engineering improvements, administrative improvements, and through the use of Personal Protective Equipment (PPE) such as safety gears. According to some studies, ergonomic hazard can be eliminated through the use of wearable sensors to detect human movement [8] and the use of predictive models [9] such as the use of machine learning techniques [10].

Improperly adjusted workstations and chairs, frequent lifting, poor posture and awkward movements, are all ergonomic hazards that are thought to cause micro trauma, which can lead to cumulative trauma disorders (CTD) and musculoskeletal disorders (MSD).

MSDs caused specifically by workplace activities are referred to as work-related musculoskeletal disorders (WRMSDs). Some examples of WRMSDs include, Tension Neck Syndrome Back Injuries, Carpal Tunnel Syndrome, Tendonitis [11].

Oftentimes, the typical worker will find themselves in a workplace environment that to a large extent lacks the congenial atmosphere that safeguards their safety and overall health. There should be a set of models that can easily predict/diagnose a worker's risk of MSD due to the symptoms shown, identify his level of severity and also proffer some solution in the ergonomics of office environments. Existing models for predicting WRMSD examined showed a high level of concern in the single classification system for present risk only. Hence, the need for a multiple classification model that can predict both present and future risk of WRMSD.

This paper focuses on the development of a multiple classification model using Decision tree and KNN algorithms to predict WRMSD. It establishes the need for the prevention of musculoskeletal disorders in corporate office workers by efficiently predicting MSD in the office environment. Subsequently, it also checks for the level of severity in a person who has been diagnosed, if it is a high or low-risk case.

2. Related Works

Sasikumar & Second (2018) created a model to estimate the likelihood of musculoskeletal hazards in computer professionals, they were able to develop a predictive model using supervised machine learning algorithms (classifiers) such as the Random Forest method and the Naive Bayes Classifier. When compared to the other algorithms, Random Forest and Naive Bayes had the highest evaluation metric values. This model was limited by small sample data size [12].

Thanathornwong et al. (2014) developed a predictive model that focused on the neck and upper-back extremities using Bayesian learning algorithm [13].

Chander & Cavatorta (2017) carried out research and developed an ergonomic risk assessment method called Postural Ergonomic Risk Assessment (PERA), using the European Assembly Worksheet (EAWS), an assessment method to assess cyclic work holistically [14] [15].

Diego-mas & Alcaide-marzal (2013) looked into systems that provide real-time feedback to the worker concerning their current ergonomic behavior with the Microsoft Kinect, the goal was to create a model that would deliver real-time feedback so that workers could correct problematic postures in real-time while on the job [16].

In a review by Ali (2016), a predictive model to identify caregivers who are at the risk of musculoskeletal disorders was addressed [17].

Suárez et al. (2014) applied the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders [10].

Jagadish & Qutubuddin (2018) carried out a study on small scale industry using Rapid Upper Limb Assessment (RULA) and Rapid Entire Body Assessment (REBA) [18].

Ribeiro et al. (2017) conducted a study on Nurses using the Portuguese version of the Nordic Musculoskeletal Questionnaire (NMQ) [19].

Panat et al. (2017) in Agricultural setting carried out a cross-sectional questionnaire survey, response rate was 90% with 88% reporting low back pain. Multiple classification was carried out [20].

Dagne et al. (2020) focused on Bank Workers, his model was based on Self-administered standard Nordic questionnaires and Multivariable binary logistic regression analyse [21].

Abledu et al. (2014) used a univariate logistic regression model to obtain estimates of the prevalence odds ratio (POR) of independent factors associated with the prevalence of WMSDs among drivers in Ghana [22].

The proposed model is an enhancement of the previously existing models. The proposed model is different from the existing models because it is a multiple classification model that does not stop at prediction alone but also goes further into checking the level of severity of those at risk, and also probes further for advanced prediction for future risk of WRMSD. Here, we used Decision tree classification algorithm and k-nearest neighbour algorithm to build the proposed predictive model for WRMSD in order to gather useful information of the dataset for accurate predictions of WRMSD amongst computer professionals in an office environment.

3. Methodology

3.1. Datasets

The dataset consists of 1,000 records comprising attributes. The dataset was generated using Macaroon online dataset generator (https://www.mockaroo.com/).

3.2. Dataset Preprocessing and Feature Selection

In the dataset the content of some records were changed to ones (1s) and Zeros (0s) for machine learning purposes. ID, weight, height, age, work duration etc would also suffice as attribute classes. Before making any actual predictions, it is always a good practice to scale the features so that all of them can be uniformly evaluated. Two new datasets were derived from the original dataset. The first dataset was used for the prediagnosis phase while the other was used for the advance diagnosis. The first dataset contained the following Features: Weight, Height, Age, BMI, Back pain, Neck pain Kneepain, Waistpain, Shoulderpain. While the second dataset contained the following features; Weight, Height, Age, days_a_week, Work duration, Standing hours, Sitting hours, and Gender.

The dataset is also divided into training and testing datasets. 80% for training and 20% for testing which gives us a better idea of how our developed model will perform during the testing phase. This way our model is tested using unseen data, as it would be in the production environment.

3.3. Process Flow of the Proposed Model

a. The proposed model will use decision tree algorithm for the preprocessing analysis of some key features in the first stage Classification.
b. After a careful preprocessing stage 1 of the dataset, the model will further be validated using the KNN algorithm to properly extract value and ensure zero errors in prediction.
c. The output of the WRMSD diagnosis will be compared to the existing models using datasets in the database to show how best the proposed system works for risk prediction and diagnosis.
d. Finally, the system generates the reports of the predictions as feedback for the user that can be sent to a printer or saved as pdf.

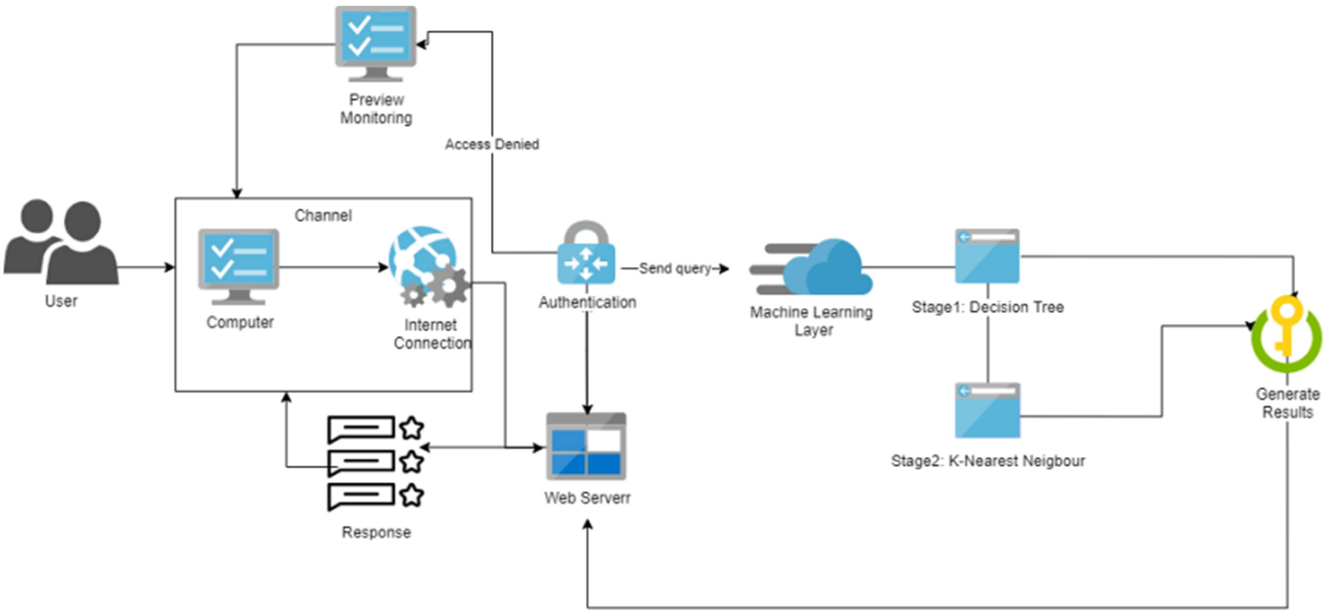

Figure 1. *The Architecture of the Proposed Model.*

Figure 1, shows the system architectural design of the proposed model. The architectural design shows the interaction between the user and the components of the system. The user uses the computer to connect to the internet. The user then sends a request which passes through the webserver to the authentication module. The authentication modules checks if the user is authorized or not. If the user is authorized the request goes through the machine learning modules where it passes through two stages of classification. The initial stage uses Decision tree classifier to determine if the user has work related musculoskeletal disorder or not and how severe it is if the user has. The second stage uses K-Nearest Neighbor classifier to determine if the user is exposed to risk of suffering from work related musculoskeletal disorder in the future. After the second stage the user generates the reports of the predictions as feedback for the user that can be sent to a printer or saved as pdf.

3.4. Model Specification

Algorithms

The different algorithms used in developing the system, includes:

(i). K-Nearest Neighbor Classifier Algorithm

$$d(x,x') = \sqrt{(x_1 - x_1')^2 + \cdots + (x_n - x_n')^2} \quad (1)$$

Equation 1; Model equation for KNN algorithm.

This classifier is referred to as the KNN Classifier for convenience. KNN addresses pattern recognition problems and is also one of the better options for tackling several classification-related jobs.

The most basic form of the K-nearest neighbor classifier algorithms predicts the target label by locating the nearest neighbor class. Distance measurements such as Euclidean distance will be used to identify the closest class.

Let (X_i, C_i) where i = 1, 2……., n be data points. X_i denotes feature values & C_i denotes labels for X_i for each i.

Assuming the number of classes as ‘c’ $_{Ci} \in \{1, 2, 3\ldots c\}$ for all values of i

Let x be a point for which the label is unknown, and we want to use k-nearest neighbor techniques to find the label class.

KNN Algorithm Pseudo code:

1. Calculate “$d(x, x_i)$” i =1, 2… n; where d denotes the Euclidean distance between the points.
2. Arrange the calculated n Euclidean distances in non-decreasing order.
3. Let k be a +ve integer, take the first k distances from this sorted list.
4. Find those k-points corresponding to these k-distances.
5. Let k_i denotes the number of points belonging to the i^{th} class among k points i.e. $k \geq 0$
6. If $k_i > k_j$ $\forall$ $i \neq j$ then put x in class i

(ii). Decision Tree Classifier Algorithm

Decision Tree is a supervised learning technique that can be used to solve classification and regression problems, however it is most commonly employed to solve classification problems. It is a tree-structured classifier in which internal nodes contain dataset attributes, branches represent decision rules, and each leaf node represents the result.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the tests are performed on the basis of features of the given dataset.

3.5. Evaluation Metrics

For evaluating the model algorithm, confusion matrix, precision, recall and f1 score are the most commonly used metrics.

3.5.1. Confusion Matrix

There is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class.

3.5.2. Accuracy

This can be described as the proportion of true results among the total number of cases examined. Used mostly for binary or multiclass classification problems.

$$\text{Accuracy}= (TP + TN)|TP + FP + FN + TN) \quad (2)$$

3.5.3. Precision

Precision is a valid choice of evaluation metric when we want to be very sure of our prediction. It helps in answering the question; what proportion of predicted positives is truly positive?

$$\text{Precision} =(TP)|TP + FP) \quad (3)$$

3.5.4. Recall

It is a valid choice of evaluation metric when we want to capture as many positives as possible. Recall is 1 if we predict 1 for all examples.

F1 score: Equally used as an evaluation metric as the harmonic mean of precision and recall.

F1 = 2*precision * recall÷ precision * recall.

4. Results and Discussion

4.1. Summary of Algorithm Analysis of the Model

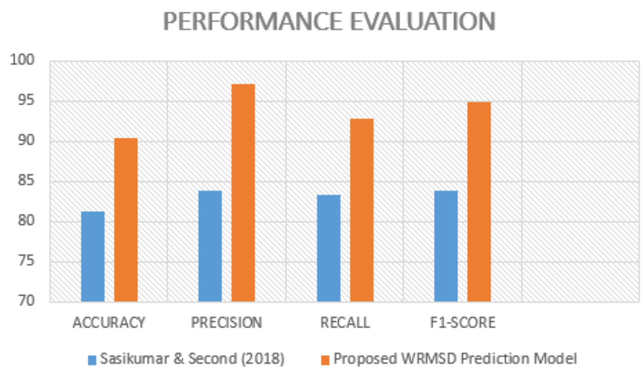

Figure 2. *Performance evaluation of proposed model against Sasikumar & Second (2018).*

Table 1. *Algorithm analysis of proposed model.*

Algorithm	Accuracy	Precision	Recall	F1
Decision Tree	96.55%	94.34%	91.45%	93.62%
KNN	90.40%	97.16%	92.71%	94.88%

Table 2. *KNN Confusion Matrix.*

KNN Confusion Matrix			
	Actual Positive	**Actual Negatives**	**Result (%)**
	True Positive (TP)	False Positive (FP)	
Predicted Positive	445	13	
Predicted Negatives	35	7	
	False Positive (FN)	True Negative (TN)	

Table 3. *Decision Tree Confusion Matrix.*

Decision Tree Confusion Matrix			
	Actual Positive	Actual Negatives	Result (%)
	True Positive (TP)	False Positive (FP)	
Predicted Positive	458	0	
Predicted Negatives	0	42	
	False Positive (FN)	True Negative (TN)	

Table 4. *Performance evaluation of proposed model against Sasikumar & Second (2018).*

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Sasikumar & Second (2018)	81.25	83.81	83.33	83.87
Proposed WRMSD sPrediction model	90.4	97.16	92.71	94.88

Table 5. *Performance evaluation of proposed model against other existing models.*

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Sasikumar & Second (2018)	81.25	83.81	83.33	83.87
Thanathornwong et al., (2014)	71.20	0	0	0
Suárez et al., (2014)	86.79	0	0	0
Diego-mas & Alcaide-marzal (2013)	88.4	0	0	0
Ali (2016)	87.2	0	0	0
Abledu et al. (2014)	55.2	0	0	0
Dagne et al. (2020)	71.9	0	0	0
Panat et al. (2017)	83.5	0	0	0
Proposed WRMSD Prediction model	90.4	97.16	92.71	94.88

Figure 2 shows the performance evaluation of proposed WRMSD model against the accuracy, precision, recall and F1-score of Sasikumar & Second (2018).

Table 5 shows the performance evaluation of proposed WRMSD model against the accuracy, precision, recall and F1-score of other existing models.

4.2. Summary of User's Response Analysis

Figure 3 depicts a pie chart representing a percentage on the 5 point rating scale. 86% of the 55 valid users strongly agrees with the developed model, 8% of the users just agrees with the model, 5% are indifferent about the model, 1% disagree with the model, while 0% strongly disagree with the WRMSD model. This acceptability test revealed that the users were satisfied with the developed model based on its effectiveness.

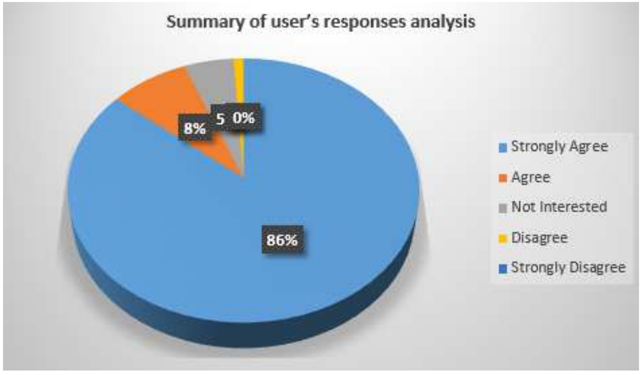

Figure 3. *User Response Analysis.*

S/n	Criteria	Strongly Agree	Agree	Not Interested	Disagree	Strongly Disagree
1.	Ease of usage	45	05	05		
2	User friendly Interface	50	03	02		
3	Speed of prediction	45	05	05		
4	Accessibility	48	02	05		
5	Effective result generation module	40	08	05		
6	Effectiveness of the diagnosis detection mechanism	50	05			

Figure 4. *Summary of user responses.*

Figure 4 depicts the system acceptability test, a questionnaire comprising of six (6) direct questions like ease of usage, efficient user interface, speed of prediction, Accessibility, effective result generation, efficiency of the diagnosis predictor, on a five (5) point Likert scale: strongly agree, agree, not interested, disagree, strongly disagree. The questionnaire was given to sixty (60) users of the predictive model to check the usability and acceptability of the classifier model. Sixty (60) responses were collected, of which 55 were valid and 5 were invalid due to users' selection of multiple entries or responses which is unacceptable for the analysis and evaluation.

5. Conclusion

This system is run over the internet as a web-based application whereby users can log in and get tested or diagnosed. To ensure the security of user details, MD5 algorithm is deployed to secure every user's password and login details from outsider attacks.

A two-step predictive model is introduced here using KNN and Decision tree machine learning algorithms. This model for predicting WRMSD enables for early detection and correction of upper and lower back disorders, carpal tunnel syndrome and other WRMSD disorders associated with office workers. Thus, potentially contributing to reduce the risk of injury due to inappropriate posture, long working hours and poorly designed workstations in corporate office workers. The knowledge gathered from this research work will also help computer using office workers reduce poor performance by workers and save a lot of cost that would have gone to WRMSD compensation and treatment.

The model also enables for early detection and correction of (WRMSD) disorders associated with office workers. Existing models for predicting WRMSD show a high level of concern in a single classification. The proposed model, however, makes it easy for multiple classifications in other to predict both present and future risk of WRMSD. Performance is estimated to have high accuracy, recall, precision and f1 score in comparison to other existing algorithms.

6. Future Work

Future research work should consider using recommender systems as more research can go into developing the model in such a way that it can proffer a solution to the user. This ability is centered on the workstation and recommending a more suitable workstation for a person at risk.

Working on further prediction which will show the exact time a person is likely to suffer from WRMSD in the near future.

Also a step can be made to work on showing the exact body part most likely to be diagnosed in the future based on the worker's earlier inputs.

This research will also benefit more from using more advanced or hybrid algorithms such as deep learning or artificial neural networks (ANN), combining decision tree classifier with genetic algorithms in carrying out the training and testing of datasets.

References

[1] Amare, D. (2010). Practice of office ergonomics on the performance of employee.

[2] Sataloff, R. T., Johns, M. M., & Kost, K. M. (2010). Practical and easy-to-implement solutions for improving safety, health and working conditions.

[3] McCauley-Bush, P. (2011). Ergonomics: Foundational Principles, Applications, and Technologies.

[4] Gupta, G., Gupta, A., Mohammed, T., & Bansal, N. (2014). Ergonomics in Dentistry. International Journal of Clinical Pediatric Dentistry, 7 (1), 30–34. https://doi.org/10.5005/jp-journals-10005-1229

[5] Kirkhorn, S. R., Earle-Richardson, G., & Banks, R. J. (2010). Ergonomic risks and musculoskeletal disorders in production agriculture: Recommendations for effective research to practice. Journal of Agromedicine, 15 (3), 281–299. https://doi.org/10.1080/1059924X.2010.488618

[6] Pascal, S. A., & Naqvi, S. (2008). An investigation of ergonomics analysis tools used in industry in the identification of work-related musculoskeletal disorders. International Journal of Occupational Safety and Ergonomics, 14 (2), 237–245. https://doi.org/10.1080/10803548.2008.11076755

[7] Okezue, O. C., Anamezie, T. H., John, J. N., & John, D. O. (2020). Work-Related Musculoskeletal Disorders among Office Workers in Higher Education Institutions: A Cross-Sectional Study. Ethiopian Journal of Health Sciences, 30 (5).

[8] Nath, N. D., Chaspari, T., & Behzadan, A. H. (2018). Automated ergonomic risk monitoring using body-mounted sensors and machine learning. Advanced Engineering Informatics, 38, 514–526. https://doi.org/10.1016/j.aei.2018.08.020

[9] Chandna, P., Deswal, S., & Pal, M. (2010). Semi-Supervised Learning Based Prediction of Musculoskeletal Disorder Risk. In Journal of Industrial and Systems Engineering (Vol. 3, Issue 4).

[10] Suárez, S.. A., Iglesias-Rodríguez, F. J., Riesgo, F.. P., & de Cos Juez, F. J. (2014). Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. International Journal of Industrial Ergonomics, 52, 92–99. https://doi.org/10.1016/j.ergon.2015.09.012

[11] Isabel, L. N. (2012). Ergonomics: A Systems Approach. Intech Open Access Publisher.

[12] Sasikumar, V., & Second, C. (2018). A model for predicting the risk of musculoskeletal disorders among computer professionals. International Journal of Occupational Safety and Ergnomics, 0 (0), 1–34. https://doi.org/10.1080/10803548.2018.1480583

[13] Thanathornwong, B., Suebnukarn, S., Songpaisan, Y., & Ouivirach, K. (2014). Computer Methods in Biomechanics and Biomedical Engineering A system for predicting and preventing work-related musculoskeletal disorders among dentists. Computer Methods in Biomechanics and Biomedical Engineering, 17 (2), 177–185. https://doi.org/10.1080/10255842.2012.672565

[14] Chander, D. S., & Cavatorta, M. P. (2017). International Journal of Industrial Ergonomics. International Journal of Industrial Ergonomics, 57, 32–41. https://doi.org/10.1016/j.ergon.2016.11.007

[15] Schaub, K., Caragnano, G., & Bruder, R. (2012). \The European Assembly Worksheet. January. https://doi.org/10.1080/1463922X.2012.678283

[16] Diego-mas, J. A., & Alcaide-marzal, J. (2013). Using Kinect Ô sensor in observational methods for assessing postures at work. Applied Ergonomics.

[17] Ali, A. (2016). A Predictive Model To Identify Caregivers At Risk Of Musculoskeletal Disorders. Electronic Theses and Dissertations, 0 (0), 2004–2019.

[18] Jagadish, R., & Qutubuddin, S.. (2018). Ergonomic Risk Assessment of Working Postures in Small Scale Industries Ergonomic Risk Assessment of Working Postures in Small Scale Industries. Grenze International.

[19] Ribeiro, T., Serranheira, F., & Loureiro, H. (2017). Work related musculoskeletal disorders in primary health care nurses. Applied Nursing Research, 33, 72–77. https://doi.org/10.1016/j.apnr.2016.09.003

[20] Panat, A. V, Avadhut Kulkarni, D., & Ghooi, R. (2017). Low Back Pain and Other Work Related Musculoskeletal Disorders and Choice of Treatment among Farmers in a Small Village of the Maharashtra State in India: A Self Reported Preliminary Study Using a Simple Questionnaire. International Journal of Health Sciences & Research (Www.Ijhsr.Org), 7, 11. www.ijhsr.org.

[21] Dagne, D., Abebe, S. M., & Getachew, A. (2020). Work-related musculoskeletal disorders and associated factors among bank workers in Addis Ababa, Ethiopia: A cross-sectional study. Environmental Health and Preventive Medicine, 25 (1). https://doi.org/10.1186/s12199-020-00866-5

[22] Abledu, J. K., Offei, E. B., & Abledu, G. K. (2014). Predictors of Work-Related Musculoskeletal Disorders among Commercial Minibus Drivers in Accra Metropolis, Ghana. Advances in Epidemiology, 2014, 1–5. https://doi.org/10.1155/2014/384279