

SCIENTIFIC REPORTS



OPEN

A Convenient Non-harm Cervical Spondylosis Intelligent Identity method based on Machine Learning

Nana Wang^{1,2}, Xi Huang¹, Yi Rao³, Jing Xiao³, Jiahui Lu^{1,2}, Nian Wang^{1,2} & Li Cui¹

Cervical spondylosis (CS), a most common orthopedic diseases, is mainly identified by the doctor's judgment from the clinical symptoms and cervical change provided by expensive instruments in hospital. Owing to the development of the surface electromyography (sEMG) technique and artificial intelligence, we proposed a convenient non-harm CS intelligent identify method EasiCNCSII, including the sEMG data acquisition and the CS identification. Faced with the limit testable muscles, the data acquisition method are proposed to conveniently and effectively collect data based on the tendons theory and CS etiology. Faced with high-dimension and the weak availability of the data, the 3-tier model EasiAI is developed to intelligently identify CS. The common features and new features are extracted from raw sEMG data in first tier. The EasiRF is proposed in second tier to further reduce the data dimension, improving the performance. A classification model based on gradient boosted regression tree is developed in third tier to identify CS. Compared with 4 common machine learning classification models, the EasiCNCSII achieves best performance of 91.02% in mean accuracy, 97.14% in mean sensitivity, 81.43% in mean specificity, 0.95 in mean AUC.

Cervical spondylosis (CS) is a degenerative disorder, an common disease, which affects up to two-thirds of the population in their lifetime¹⁻⁴. The CS seriously affect people's physical and mental health and quality of life and increase the burden on individuals and society. What's more, it leads to the loss of human-related functions and is accompanied by depression, anxiety and other psychological damage. The main complaint of CS is neck pain which is reported by approximate 30–50% from the patients⁵. Meanwhile, the global point prevalence of neck pain was 4.9% and the neck pain ranked 4th highest in terms of disability as measured by YLDs, and 21st in terms of overall burden in the Global Burden of Disease 2010 Study⁶. The early detection of the CS is critical for burden lighten. As the earlier the disease is discovered, the easier it is to treat, the higher the cure rate is, and the less the patient spend. The CS is a chronic 'wear and tear' degenerative process of the cervical spine that initially is the vertebral bodies and intervertebral disks degeneration in the neck, and can develop into disk ruptures and herniation, osteophyte, compression of the spinal cord, or cervical spondylotic myelopathy (the most important complication of degenerative disease of the cervical spine)⁷⁻⁹. As cervical degeneration worsens, clinical manifestations become more obvious, and the difficulty and cost of treatment increases. In detail, for the patients with the vertebral bodies and intervertebral disks degeneration in the neck, non-operative treatment continues to play an important role in treatment². For the patients with cervical spondylotic myelopathy, surgical treatment have been conventional means but may lead to significant problems including adjacent level^{2,10}.

The pathogenesis and clinical symptoms of the CS are complex. For the pathogenesis, it is supported by the modern medical theory that the chronic degeneration of the cervical spine lead to the CS. The degeneration cause the changes in the morphological or structural of the spine, such as loss of disc height, disk ruptures and herniation, narrowing of the foramina, osteophyte. Since the morphological structure of the spine can be records by imaging instruments¹¹⁻¹⁶, the change above can be directly observed from the imaging information. For example, the oblique radiograph of the cervical spine in a patient with CS shows the loss of disc height, anterior osteophytosis, and narrowing of the foramina¹⁷. For the clinical symptoms, the clinical symptoms accompanying the occurrence and development of CS are complex and varied^{17,18}. In the early stage of CS, the suffers are usually harassed by the neck pain, cervical stiffness and other discomfort in the neck, shoulder blades, and upper limbs. As the disease worsens, the range of neck movement is limited, and the pain is aggravated by the movements.

¹Institute of Computing Technology(ICT), Chinese Academy of Sciences(CAS), Beijing, China. ²University of Chinese Academy of Sciences, Beijing, China. ³Xiyuan Hospital, China Academy of Chinese Medical Sciences(CACMS), Beijing, China. Correspondence and requests for materials should be addressed to L.C. (email: lcui@ict.ac.cn)

Due to the complex above, the identification of CS is a sophisticated and complicated work. The classification of CS, the clinical symptoms and imaging characteristics are presented in the CS diagnostic criteria. Combined with the professional knowledge as well as clinical experience, the doctor collects relevant symptom information in the form of question and answer, analyze the symptoms and imaging information, and gives a diagnosis result according to the diagnostic criteria. However, when diagnosed by clinical symptoms, the accuracy of the diagnosis result is hindered by the asymptomatic, slight early symptoms that are easily overlooked, individual differences, and unrelated causes (such as the neurological disease). When diagnosed by observing the morphological or structural changes of the spine, the cost of the diagnosis is increased by the expensive medical instrument. When diagnosed by the clinical approach, the timeliness and cost of the diagnosis is affected by the cumbersome diagnostic process, which involves medical instruments resource and need the intervention of doctors or experts. What's more, the concern on the healthy will be aggravated by frequent inspection of medical instruments, for example the radiation.

It is also agreed by the tendons injury theory of traditional chinese medicine that the pathological changes of the neck muscle system (the tendons of neck) is the causative factor of CS. In detail, the pathological changes of the cervical muscle system leads to the decline in its own mechanical properties, the destruction in the exogenous stability of the cervical vertebrae, aggravate the degeneration of the cervical spine, and finally form the CS¹⁹. The pathological changes of the neck muscle system and cervical spine cause chronic dysfunction as well as pain²⁰, leading to anomalous pattern of muscle activity^{21–27}. And when the muscles are activated in the activity, the Motor Unit Action Potential Trains (MUAPTs) are generated by motor units, superimposed on the surface of the skin and form a non-stationary weak signal which can be acquired by the sEMG device and generate electromyography. Therefore the pathological changes can be captured by the sEMG device. Meanwhile, there are many researches on using myoelectricity to study the muscle activity and functional status^{21–27}, which demonstrates that there are differences in sEMG signals between population with cervical musculoskeletal disorders or neck pain and the healthy. What's more, the tender point refers to the area where pain can occur when pressed, the distribution of which is closely related to biomechanics²⁸. It is the signs of neck muscle pathological changes^{19,29}. The tender points of patients with CS mostly located on the cervical paravertebral muscle, trapezius and sternocleidomastoid muscle³⁰. Thus, it provides a chance that we can explore the relationship between sEMG signals on the muscles of the cervical paravertebral muscle, trapezius and sternocleidomastoid muscle and CS to identify CS. Benefit from the development of sensors technology, sEMG device become more portable and more cheaper, promoting the sEMG technology to become a competitive choice for the convenient CS identification. The sEMG have attracted a lot of attention in muscle function assessment^{31,32}, muscle activity assessment^{23,24,26}, rehabilitation effect tracking³³ and rehabilitation guidance³⁴.

Meanwhile, the statistical analysis also opens a window onto wellness³⁵. The traditional statistical analysis methods, has been widely used in clinical data analysis, for instance linear or logistic regression^{36–38}. As the increasing data become larger and complicated, modern statistical methods have been used to deal with the complicated data³⁹. The machine learning, a powerful analysis methods, has made great progress in medical^{40–45}. With the convenient of sEMG technology and the development of artificial intelligence, the convenient, no-harm, intelligent CS identification method can be considered. The data acquired by portable sEMG device present a huge challenge to the identification of CS. The high-dimensional sEMG data can cause dimensional disaster which decrease computational efficiency, increase memory storage requirements, and cause overfit. Faced with the high-dimensional data, feature extraction and feature selection, which are effective means of data preprocessing, have the advantages of improving model performance, increasing computational efficiency, decreasing memory storage requirements, giving model better readability and interpretability, and building better generalization model⁴⁶. Besides, the data shows weak availability of faulty, redundant, insufficient, sparse distribution since the data acquisition is susceptible to external factors in non-lab environments by portable sEMG device. So a powerful machine learning model should be considered. Ensemble learning containing a number of weak learners can not only learn linear and complex nonlinear function but also boost weak learners which are slightly better than random guess to strong learners which can make very accurate predictions⁴⁷. The gradient boosted regression tree (GBRT), one of powerful ensemble learning, has been successfully used in classification task^{48–51}. And it is a competitive choice for the classification task on limited weakly available data.

In this work, we proposed a new convenient, non-harm and intelligent method EasiCNCSII to identify CS based on sEMG and machine learning as the Fig. 1 shown. The method mainly consist of data acquisition and CS identification. For data acquisition, we proposed a convenient, time-saving data acquisition method, which involves 6 muscles and 7 movements (see Supplementary for the selection of muscles and movements). The user only need to spend less than 20 minutes independently completing a set of simple movements according to instruction, after connecting the portable sEMG device to the laptop and user's neck muscles. The relevant data is uploaded to the intelligent processing terminal while being collected by the sEMG device. For CS identification, the EasiAI model based on 3-tier architecture, was developed to identify CS. The EasiAI consists of feature extraction, feature selection and classification algorithm. For feature extraction, we extract 11 types of features from raw sEMG signal, of which 6 types are extracted in the common high dimensional time series feature extraction methods, such as time-domain method, of which 5 types are built inspired by the relevant knowledge. Most of the features are proved to be significantly associated with the CS by Pearson ($p \leq 0.05$). For feature selection, EasiRF, a feature selection method, was developed to select the most relevant features and improve the performance of the CS identification. The easyRF is validated effective compared with traditional feature selection algorithms. For classification algorithm, a classification algorithm based on GBRT is developed to identify CS. The EasiAI achieve the best performance with 91.02% in accuracy, 97.14% in sensitivity, and 81.43% in specificity compared with 4 kinds of machine learning classification model. The EasiCNCSII is validated effective.

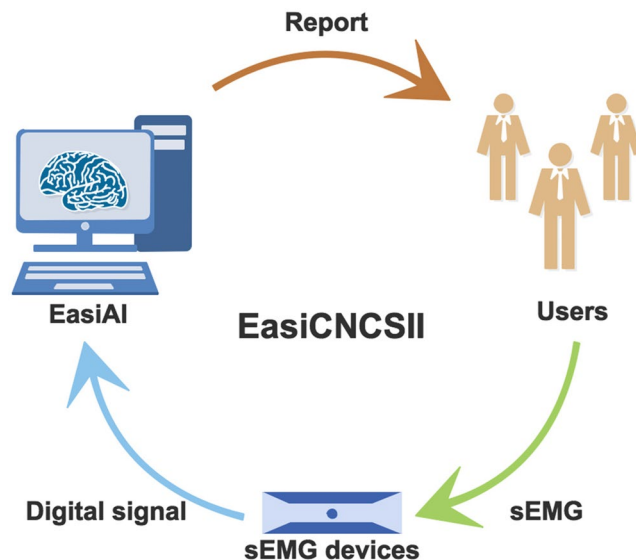


Figure 1. The CS identify based on sEMG and machine learning. Firstly, users perform a set of movements according to instructions. Secondly, the portable sEMG device acquire sEMG signals of the user and send it to the smart terminal. Finally, the EasiAI, an intelligent CS identification model, predict the state of user using the sEMG signals and return the report which shows whether the user suffer from the CS.

Methods

Participates. The 179 volunteers participated in the study from March 15, 2017 to December 31, 2017 in China, of which 109 were CS suffers and 70 were the healthy. The former have received a clinical diagnosis of the CS, which are in accordance with 2012 ICD-9-CM Diagnosis Code 721 (721.0 Cervical spondylosis without myelopathy, 721.1 Cervical spondylosis with myelopathy) and the criteria of diagnosis and treatment for CS issued by China Rehabilitation Medicine association. The latter free from the CS are diagnosed by the rich experienced clinicians. The exclusion criteria includes cervical vertebral trauma, cervical vertebral surgery, congenital spinal deformity, syringomyelia, amyotrophic lateral sclerosis, spinal cord tumor, spinal cord injury, adhesive arachnoiditis, the fasciitis, cervical injury, tumor or infection, pregnant, breast-feeding, menstruation. All data collection protocols were approved by Institute of Computing Technology, Chinese Academy of Sciences. Written informed consents were obtained from all volunteers. The acquisition method was carried out in accordance with the approved guidelines. See Supplementary for the instruction of data acquisition.

The EasiCNCsII. The EasiCNCsII is designed as shown in Fig. 2, which includes two parts: the data acquisition and the EasiAI model. For data acquisition, the users, after connected sEMG device with the 6 muscles, complete 7 movements according to the simple instructions. The analog signals from the users are converted into the digital signals, which are high dimensional time series data, and sent to the EasiAI by the sEMG device. The EasiAI is a 3-tiers data processing model: feature extraction, feature selection and classification algorithm. Feature extraction algorithm extracts features from the a user's digital signals in the methods of time-domain, frequency-domain, time-frequency-domain, etc. Feature selection algorithm EasiRF based on the Random Forest (RF) is developed to selected the most relevant features. The classification algorithm based on the gradient boosted regression tree is developed to identify CS, achieving the good performance on limited data set with relative small computing overhead. Using the input data after feature selection, the report, which show whether the user suffers from the CS, is generated and returned to the user. The lightweight algorithms can be integrated into the user end and quickly send report to users without concerns on privacy.

Feature extraction. The 5 common methods: time-domain, frequency-domain, time-frequency, parametric model and nonlinear feature analysis are used to extracted features from the sEMG signal which are high-dimensional time series data. Besides, the method of using the disease-related knowledge to construct features are also considered. With the 5 common methods above, we extracted 63 features from S_{ij} , of which 11 are extracted in methods of time-domain^{52,53}, of which 14 are computed in methods of frequency-domain^{52,53}, of which 23 are computed in methods of time-frequency based on wavelet transform and wavelet packet transform^{54,55}, 14 of which are extracted in AR parametric model⁵⁶, of which 1 nonlinear entropy feature^{52,53,57} are extracted. Among the 63 features above, the 5 features including the root mean square (RMS), median frequency (MF), mean power frequency (MPF), the average electromyogram (AEMG), and the integrated electromyogram (IEMG) are common features in clinical research. Besides, considering the relative knowledge of the CS, the 45 features are extracted from S_p , of which 2 called as cervical flexion-relaxation ratio (FRR)^{58,59} is commonly used in clinical research and can only be extracted from the S_p , and of which 43 are new features. The 63 features from S_{ij} and 45 features from S_i are divided into 11 types: TF, FF, WL, WLP, AR, EY, FRR, DU, ACI, UN, SYM, facilitating statistical analysis. Details on the calculation are shown in the feature extraction part of the Supplementary and Supplementary Tables S1 and S2.

The EasiRF algorithm. The EasiRF based on the RF is a stochastic model since the samples are randomly selected to generate the tree and the features are randomly selected to be splitting rule. Firstly, we generate 7 different data sets and iteratively use RF to train on each data set. Secondly, set the tree number of RF model different, select the top 25 most important features and merge the features of each iteration until the feature number of the merged set is not growing in each iteration. Finally, the final feature set is generated by merging 7 feature sets from 7 data sets above. As shown in Algorithm 1, the feature selection algorithm EasiRF based on RF is developed to get the most relative features.

The classification algorithm. A classification algorithm based on Gradient boosted regression tree (GBRT) is developed to identify CS. The GBRT is also known as gradient boosting machine (GBM) or gradient

Algorithm 1. The EasiRF algorithm.

Input: the data set S

Output: the set of features $FeaSet$

- 1: $k=7, FeaSet = \emptyset$
- 2: Divide the S into S_1, S_2, S_3, \dots , and S_7
- 3: **for** $j = 1$ to 7 **do**
- 4: $DS = S - S_j$
- 5: $FeaSet_j = \emptyset$
- 6: $Num = 0$
- 7: **while** $|FeaSet_j|$ is not growing **do**
- 8: $FeaSet_t = \emptyset$
- 9: Set the number of the RF with $(50 + Num * 5)$.
- 10: Run the RF with DS , output the features and its importances.
 Put the 25 the most important features into $FeaSet_t$
- 11: $FeaSet_j = FeaSet_j \cup FeaSet_t$
- 12: $Num = Num + 1$
- 13: **end while**
- 14: $FeaSet = FeaSet \cup FeaSet_j$
- 15: **end for**
- 16: **return** $FeaSet$

tree boosting⁶⁰. It is a tree ensemble model and its base classifier is a decision tree that can be used both for regression and for classification. The GBRT model is trained iteratively. For each iteration, the sample instance to minimize the objective is added. In detail, if a given dataset D ($D = (x_i, y_i), 1 \leq i \leq n, x_i \in R^m, y_i \in R$) consist of n samples with m features, and $\hat{y}_i^{(t)}$ is the prediction of the i -th instance at the t -th iteration, the GBRT can be trained after t iteration, for each of which we add f_t with the minimum objective function $L^{(t)}$ value in Formula 1.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) + g(f_t) \quad (1)$$

Here, the $L^{(t)}$ is the objective function in t -th iteration.

The metric of model performance. The accuracy, sensitivity, specificity, FNR, FPR and AUC (the area under the sensitivity and specificity curve) are used to evaluate the performance of algorithms in our work in Formula 2–6. The large the accuracy, sensitivity and specificity are, the smaller the FNR and FPR are, the better performance of the model is, the more efficient the feature selection algorithm is. Besides, the larger AUC value indicates higher classification accuracy across a range of threshold choices^{40,61}, so the AUC is also used to illustrate the performance of the classifier.

$$\text{accuracy} = \frac{\text{true positive} + \text{true negative}}{\text{positive} + \text{negative}} \quad (2)$$

$$\text{sensitivity} = \frac{\text{true positive}}{\text{positive}} \quad (3)$$

$$\text{specificity} = \frac{\text{true negative}}{\text{negative}} \quad (4)$$

$$\text{FNR} = \frac{\text{false negative}}{\text{positive}} \quad (5)$$

$$\text{FPR} = \frac{\text{false positive}}{\text{negative}} \quad (6)$$

The ‘true positive’ is the number of correctly predicted patient with CS, ‘positive’ is the number of patient with CS shown, ‘true negative’ is the number of correctly predicted healthy people free of CS, the ‘negative’ is the number of healthy free of CS shown, the FNR is false negative rate, the FPR is false positive rate, ‘false positive’ is the number of predicted patient with CS which are actually the healthy, and ‘false negative’ is the number of predicted healthy people which are actually the patients with CS. The higher the accuracy, sensitivity and specificity are, the better the performance of model is. The lower the false positive rate and false negative rate are, the better the performance of model is.

Handling of missing data. Some sample data were missing for some variables. We filled in missing data with the mean value. In detail, the missing data from CS suffers are filled with the mean values of all the samples from the CS suffers. The missing data from health are filled with the mean values of all the samples from the healthy.

Experimental set-up. The python (version 2.7.13) and matlab (version 2016r) were used to implement features extraction from sEMG data. The python (version 3.6.0) and xgboost (version 0.6) were used to implement the EasiAI model. The RF, NB, LR and SVM model are implemented by the scikit-learn (0.19.1) and the python (version 3.6.0). The laptop and sEMG device which used to collect the sEMG signal are provided by Wireless Sensor Network Lab (this research does not develop the hardware). The EasiAI is deployed in the laptop with the CPU of i7, the memory of 8GB and operating system of 64-bit.

Code availability. After we reorganize the codes of the EasiAI, the source codes will be available at github. Currently, it is available from wangnana@ict.ac.cn on reasonable request.

Ethical approval. The study was approved by Institute of Computing Technology, Chinese Academy of Sciences. All volunteers signed informed consents. The acquisition method was carried out in accordance with the approved guidelines.

Result

Data preparation. The dataset are made up of 537 samples from 179 volunteers from March 15, 2017 to December 31, 2017 in China, including 109 CS suffers and 70 healthy population. Each volunteer’s sEMG signal was acquired repeatedly 3 times and form 3 samples as shown in Supplementary Fig. S1 (see Supplementary for the instruction of data acquisition). We split the dataset into training samples (training set), validation samples (validation set), and test samples (test set) according to 16: 4: 5. Meanwhile, it is ensured that samples belonging to the same volunteer are only distributed in one of the three sets above. Each sample consists of 7 multiple

	Accuracy	Sensitivity	Specificity	FNR	FPR
Non	86.54%	92.51%	77.14%	7.49%	22.86%
FS	87.10%	95.28%	74.29%	4.72%	25.71%
CIFE	86.00%	92.60%	75.71%	7.40%	24.29%
MCFS	83.16%	92.60%	68.57%	7.45%	31.43%
F-score	87.10%	95.28%	74.29%	4.72%	25.71%
EasiRF	91.02%	97.14%	81.43%	2.86%	18.57%

Table 1. The comparison of performance with different feature selection algorithms.

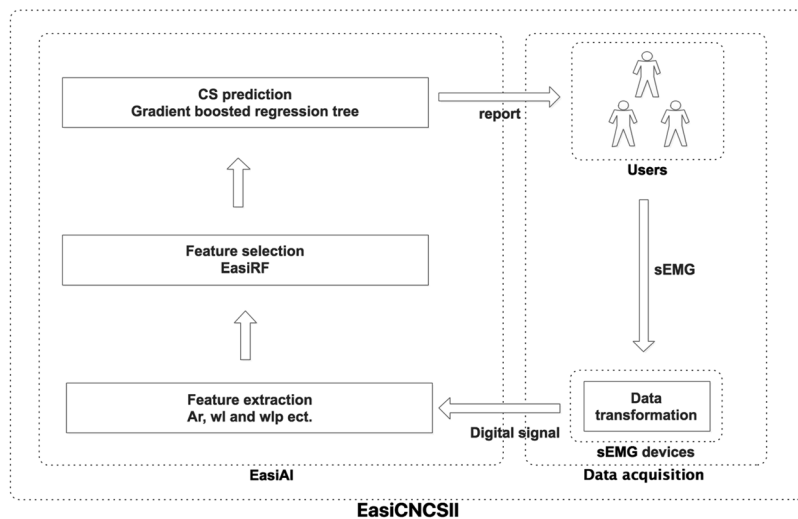


Figure 2. The EasiCNCsII method. The left side of the figure is the CS identify model EasiAI consisting of feature extraction, feature selection, classification algorithm. The right side of the figure is data acquisition. The data collected by sEMG is automatically transmitted to the intelligent terminal equipped with the EasiAI. The report generated by the EasiAI is sent to the users.

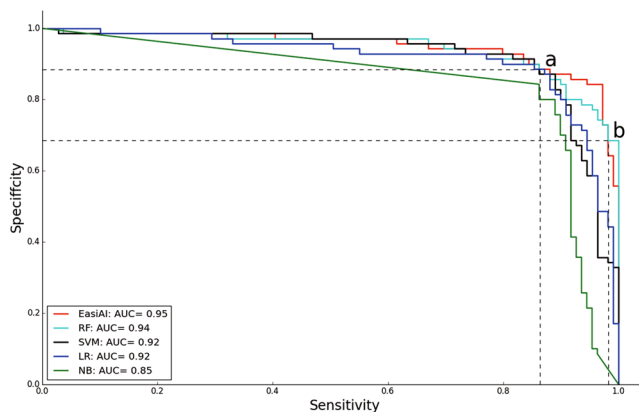


Figure 3. The performances of the EasiAI, RF, SVM, LR, and NB. The x-axis is the sensitivity. The y-axis is the specificity. Different color denotes different machine learning model. The curves of different colors consist of sensitivity and specificity points of different models. The closer the sensitivity and specificity points are to the upper right corner, the larger the sensitivity and specificity are, the better the model performs. All the sensitivity and specificity points of the NB are the farthest from the upper right corner and fall below the curves of other models. The two curves of the LR and the SVM are intertwined whose most of sensitivity and specificity points fall below the curve of the RF. The two curves of RF and EasiAI are intertwined and most of the RF's sensitivity and specificity points fall below the ones of the EasiAI, especially between point a and b. The sensitivity and specificity points of the EasiAI are closest to the upper right corner, and have the highest AUC value 0.95.

high-dimensional time series data from a sample, each of which is generated from the movement A_i (the A_i includes bow (A_1), head backwards (A_2), left flexion (A_3), right flexion (A_4), left rotation (A_5), right rotation (A_6), and hands up (A_7)) and represented as S_i . The S_i is expressed as Equation 7. The $S_{i,j}$ denotes the sEMG signals

	Accuracy	Sensitivity	Specificity	FNR	FPR
EasiAI	91.02%	97.14%	81.43%	2.86%	18.57%
RF	87.65%	93.42%	78.57%	6.60%	21.43%
SVM	86.00%	89.78%	80.00%	10.22%	20.00%
LR	86.57%	88.92%	82.86%	11.08%	17.14%
NB	82.10%	89.83%	70.00%	10.17%	30.00%

Table 2. The performance of machine learning on test set.

collected from the muscles M_j (the M_j includes left sternocleidomastoid (M_1), left upper trapezius (M_2), left cervical erector spinae (M_3), right cervical erector spinae (M_4), right upper trapezius (M_5), and right sternocleidomastoid (M_6)) activated by movement A_p , expressed as Equation 8.

$$S_i = [S_{i,1}, S_{i,2}, \dots, S_{i,j}]^T, \quad (1 \leq i \leq 7, 1 \leq j \leq 6) \quad (7)$$

$$S_{i,j} = [p_1, p_2, \dots, p_n] \quad (8)$$

Here, the S_{ij} denotes the electrical signal data produced by the muscle M_j throughout the course of the movement A_i and consists of the p_n which is the value of the float type in python 2.7.

Feature extraction. For each sample, the 423 ($45 + 6 \times 63$) features are extracted from A_1 and 421 ($45 + 6 \times 63$) features (the 421 features doesn't include FRR features) are respectively extracted from the other 6 movements. Thus, 2949 ($423 + 6 \times 421$) features are extracted from the raw sEMG signal generated by all the movements above which include $A_1, A_2, A_3, A_4, A_5, A_6, A_7$. The 2949 features's distribution are shown in Supplementary Fig. S3. The Pearson correlation indicated that 1789 features are significantly associated with the CS ($p \leq 0.05$). (The significant features's distribution are also shown in Supplementary Fig. S3).

Feature Selection. We developed a feature selection algorithm EasiRF based on RF to select the most relevant features and improve the performance of the CS identification. The EasiRF divide the 537 samples into 7 different data sets, each of which is represented as D_i ($0 \leq i \leq 7$). The RF (The number of trees is set to different value in each iteration.) is iteratively used to select the top 25 most important features from D_i . And put the 25 selected features of each iteration into the feature set S_i until the feature number of the S_i is not growing. As shown in Supplementary Fig. S4, the selected feature number tends to be stable on each data set D_i when the iteration number reaches 40. The final feature set including 282 features are generated after merging all the S_i above (the feature type distribution of 282 features are shown in Supplementary Fig. S5).

In order to validate the effectiveness of the EasiRF, the Fisher Score (FS)⁶², Conditional Infomax Feature Extraction (CIFE)⁶³, Multi-Cluster Feature Selection (MCFS)⁶⁴, f-score are respectively selected from 4 kinds of traditional feature selection algorithms⁴⁶ as well as EasiRF. With the metrics of accuracy, sensitivity, specificity, FNR and FPR, we compare the performance of the GBRT on test set with the feature selection algorithms above, using five-folds cross-validation. As shown in Table 1, the mean accuracy of the model is 86.54% without feature selection. In spite of the slight drop in mean accuracy with the MCFS and CIFE, the mean accuracy of other algorithms is over 87.10%. What's more, the mean accuracy with the EasiRF is the highest with the value of 91.02%. The mean sensitivity of the model is 92.51% without feature selection. All the mean sensitivity of the model are over 92.51% with feature selection algorithms. And the mean sensitivity with the EasiRF is the highest with the value of 97.14%. The mean specificity is 77.14% without feature selection. However, the mean specificities with feature selection algorithms are less than 77.14%, except the EasiRF with the highest value of 81.43%. The mean FNR of the model is 7.49% without feature selection. All the mean FNR with the feature selection algorithms are less than 7.49%. The mean RNR with the EasiRF is the smallest with the value of 2.86%. The mean FPR of the model is 22.86% without feature selection. And the mean FPR with feature selection algorithms are larger than 22.86%. The FPR with the EasiRF are the smallest with 18.57%. Compared with the feature selection algorithms above, the EasiRF perform best and are validated most effective. The comparison of more feature selection algorithms are shown in Supplementary Table S3.

The train and test of the EasiAI. We train the EasiAI on training set and validate on validation set. The performance of the classifier is mainly affected by model parameters, especially the number of weak classifiers. Thus we first assessed that the number of trees (weak classifiers) included in the EasiAI were enough to obtain the highest AUC on validation set. As shown in Supplementary Fig. S6, the higher the number of trees is, the higher the AUC on the validation set is. Furthermore, the AUC on the validation set achieve highest when the tree number is 535. The other parameters of the EasiAI model are shown in Supplementary Table S4. The final model is generated by training EasiAI using the parameters in Supplementary Table S4 on the dataset consisting of the training set and validation set.

To validate the performance of the final EasiAI, we tested the model on test set, using five-fold cross-validation. The metrics are accuracy, sensitivity, specificity, FNR, FPR as well as AUC. The large accuracy, sensitivity, specificity and AUC are, the smaller false negative rate and false positive rate are, the better the model classification performance is. The AUC, the mean accuracy, the mean sensitivity, the mean specificity respectively are 0.95, 91.02%, 97.14% and 81.43% as shown in Fig. 3 and Table 2. The mean FNR, the rate of missed diagnosis, is 2.86%.

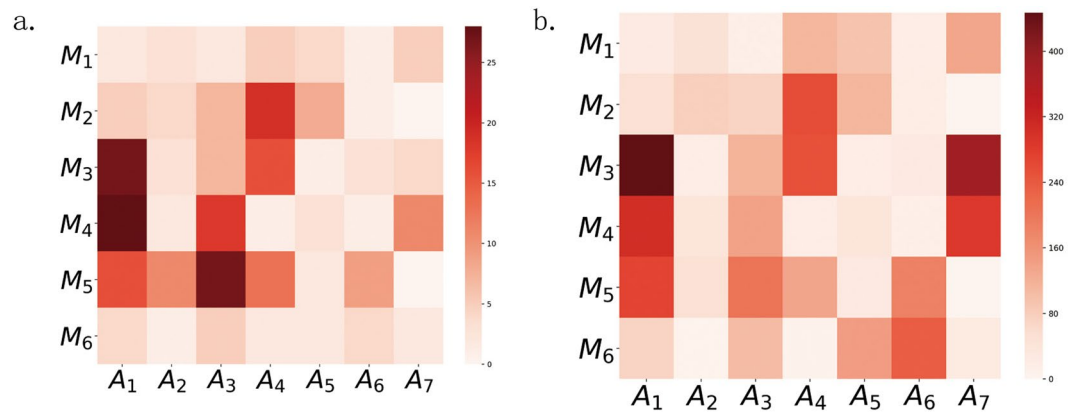


Figure 4. The heat map. The $M_1, M_2, M_3, M_4, M_5, M_6$ respectively denotes left SCM, left UT, left CE, right CE, right UT and right SCM. The $A_1, A_2, A_3, A_4, A_5, A_6$ and A_7 respectively denotes low head, head backwards, left flexion, right flexion, left rotation and right rotation, hands up. (a) The number distribution of features in muscles and movements. A square represent the number of feature on M_i activated by A_j . The darker the square's color is, the more the features on the square are. (b) The importance distribution of features in muscles and movements. A square represent the importance of feature on M_i activated by A_j . The darker the square's color is, the more important the features on the square are.

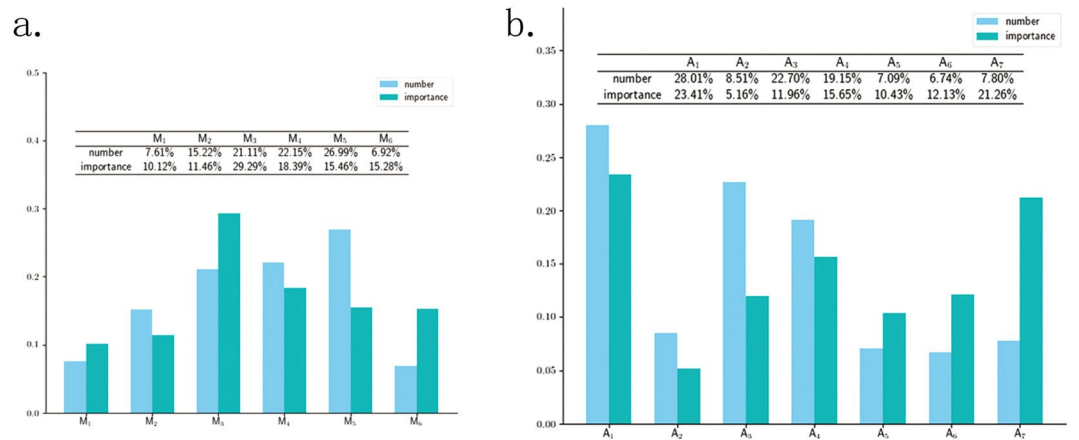


Figure 5. The feature distribution. (a) The feature distribution on the muscles. The $M_1, M_2, M_3, M_4, M_5, M_6$ respectively denotes left SCM, left UT, left CE, right CE, right UT and right SCM. The light bars show the distribution of feature numbers (A-distribution). The right UT rank first, followed by right CE, left CE, left UT, left SCM, and right SCM. The right UT, right CE and left CE take up 70.25%. The dark bars show the distribution of feature importance (B-distribution). The left CE rank first, followed by right CE, right UT, right SCM, left UT and left SCM. The left CE, right CE and right UT take up 63.14%. (b) The feature distribution on the movements. The $A_1, A_2, A_3, A_4, A_5, A_6$ and A_7 respectively denotes low head, head backwards, left flexion, right flexion, left rotation, right rotation and hands up. The light bars show the distribution of numbers (C-distribution). The movement of head backwards rank first, followed by left flexion, right flexion, low head, hand up, left rotation, right rotation. The dark bars show the distribution of feature importance (D-distribution). The movement of the bow head rank first, followed by hands up, right flexion, right rotation, left flexion, left rotation, head backwards.

The mean FPR, the misdiagnosis rates, is 18.57%. Overall, the accuracy of model is higher than 90%, the missed diagnosis rate of our model is less than 3%, and is validated effective.

Comparing with support vector machines (SVM), Logistic regression (LR), NativeBayes (NB), Random Forest (RF).

The identification of CS is a task of classification. Faced with limited data, we chose the following four common models: Support vector machines (SVM), Logistic regression (LR), Native Bayes (NB), and Random Forest (RF). We validate the effectiveness of the EasiAI by comparing the EasiAI with the SVM, LR, NB and RF, on test set (the test set only includes the 282 features which are selected from the 2949 features by the EasiRF) in the same classification task, using five-fold cross-validation. In each fold, we use the grid search approach to get the important parameter value, with which the model perform optimal. The details on the parameters of the EasiAI, SVM, LR, NB and RF are shown in the Supplementary Tables S4–S8. As shown

	Inquiry	Images examination	EasiCNCSSII
Time	dozens of minutes	dozens of minutes	about 20 minutes
Cost	low	high	middle
Harmful	no	yes	no
Location	hospital	hospital	no limite
Convenience	low	low	high
Doctor/professional	yes	yes	no
The time of getting the results	a few minutes	days	several seconds

Table 3. The comparison of the Inquiry, Images examination, EasiCNCSSII.

in Table 2, the highest mean accuracy is 91.02% achieved by the EasiAI, and the lowest mean accuracy is 82.10%. The highest mean sensitivity is 97.14% achieved by the EasiAI, and the lowest mean sensitivity is 88.92%. The highest mean specificity is 82.86% and the lowest mean specificity is 70.00%. The mean specificity of the EasiAI is 81.43%, just smaller than the highest. The highest mean FNR is 11.08% and the smallest value is 2.86% achieved by the EasiAI. The highest mean FPR is 30.00% and the lowest value is 17.14%. The mean FPR of the EasiAI is 18.57%, just larger than the smallest one. Besides, the curves of sensitivities and specificities of models above are shown in Fig. 3. Most of the sensitivity and specificity point falls below the red curve of the EasiAI, especially between point a and b. The EasiAI achieve the best performance compared with RF, SVM, LR and NB with the metrics above.

Analysis of learned knowledge about CS. The EasiAI achieved high accuracy in identifying CS, so we believe that the features in final model play an important role in the classification. And the informative knowledge about the CS can be learned from the feature distribution in muscles and movements.

The number of features and feature importance are two important attributes. The feature number is the number of features extracted from muscles or movements. The more the features that are distributed on the muscle/movement are, the stronger the muscle/movement have the ability to identify CS, the more the muscle/movement have differences between the healthy free from the CS and the CS suffer, the more the muscle/movement is related to CS. The feature importance is the contribution to the performance of task above in the training model. The more important the features from the muscle/movement are, the stronger the muscle/movement have the ability to identify CS, the more the muscle/movement have differences between the healthy free from the CS and the CS suffer, the more the muscle/movement is related to CS.

In order to further analyze the feature distribution above, the number distribution and importance distribution of the features in the muscles and movements are plotted as units of the heat map in Fig. 4. The darker the color of the unit is in Fig. 4, the more the features on the unit are in Fig. 4(a), and the more important the features on the unit are in Fig. 4(b). As shown in Fig. 4, the features on the units formed by the M_3 (left CE) and A_1 (bow), M_4 (right CE) and A_1 (bow), M_2 (left UT) and A_7 (hands up), M_5 (right UT) and A_7 (hands up) are the most in number and most important for the CS identification performance. It is concluded that the CE activated by the movement of A_1 (bow head) and the UT activated by the movement of A_7 (hands up) may be more closely related to the CS.

In order to further understand feature distribution, the feature number distribution in muscles (A-distribution), the importance distribution in muscles (B-distribution), the feature number distribution in movements (C-distribution), and the features importance distribution in movements (D-distribution), are plotted as Fig. 5. As shown in Fig. 5(a), the features on UT (M_2 and M_5) and CE (M_3 and M_4) take up 85.47% in the A-distribution and 74.60% in the B-distribution. As shown in Fig. 5(b), the features on the A_1 (bow) rank first in both C-distributions and D-distribution. The features on A_7 (hands up) rank second in D-distribution. It is concluded that the CE play an most important role in identifying CS followed by UT. And the movements of A_1 (bow head) also play an most importance role in identifying CS, followed by the movement of A_7 (hands up).

The conclusion can be explained by that the more frequently muscle are activated, the more the muscle (movement) is related to CS. Most of volunteers are sedentary population who frequently bow head and use the hands and arms to work. The movement of bowing head activate the CE, and the movement of the hands and arms activate the UT. The excessive use of the CE and UT cause muscle strain, cause dysfunction of the spine stabilizing system, and accelerate cervical degeneration and lead to CS. It may provide a suggestion for sedentary population to prevent CS by avoiding excessive use of the CE and UT, strengthen the UT and CE protection.

Discussion

The CS identification methods mainly include clinical symptoms examination and imaging examination since the diagnosis of CS is determined by clinical symptoms and imaging information^{16,17}. Currently, the clinical symptoms examination are performed by experts or doctors in the form of the inquiry. The images examination, which mostly includes spinal angiography, vertebral artery angiography, X-ray, computed to-mography (CT), and magnetic resonance imaging (MRI), etc, mainly depend on observation of the physical changes in spine and its subsidiary structure by imaging instruments to identify CS. The EasiCNCSSII depend on detecting abnormal sEMG signal associated with muscle activity to identify CS. We compare the practicality of the EasiCNCSSII with the inquiry, imaging method as shown in Table 3. The inquiry method is easiest and fastest. However, the method is suitable for population with obvious symptoms, for example severe pain, since information is mainly

determined by suffer's subjective feelings and judgment. And it needs the instruction of the doctor and auxiliary equipment to accurately identify CS. The images examination is the essential CS examination currently. However, its cost is relatively high and it is time-consuming considering time spent to go to the hospital and wait for the results which requires the intervention of doctors or experts. What's more, it can not be used frequently with concern on health since the frequent use of imaging instruments can put a strain on the health, for example the radiation. Compared with inquiry and imagines examination, the EasiCNCSII is an best choice to identify CS outside the clinic with the advantage of easy use, low cost, no-harm. Owing to the intelligent algorithm, users can get results quickly after examination. What's more, the wearable sEMG acquisition technology get more and more attention⁶⁵. Combined with mobile application, the EasiCNCSII potentially provide low-cost convenient universal access to indispensable care outside the clinic, and even promote the development of telemedicine, especially in areas short of medical resources.

To our best knowledge, previous research on CS identification based on the sEMG and machine learning are few, so there is a lot room for improvement. Traditional classification or regression algorithms can achieve good performance when fed with a wealthy of high quality data. However, the amount of data is limited and the data dimension is high. Besides, data acquisition is vulnerable to the environment so that poor quality data are often collected by portal sEMG. The EasiAI can handle these influences by boosting multiple weak learners to reach higher prediction accuracies. Compared with the RF, SVM, LR and NB, the EasiAI achieves best performance, and identifies complex crowd with low missed diagnosis rate. It is hard to fully understand the relation between CS and the sEMG signal from the activity of the deep and shallow muscles with the limited data, but the data-driven machining learning can achieve better performance and accelerate our understanding of the principles behind the CS identification to assist diagnosis and guide treatment as more data are accumulated. We are looking forward to more convenient and intelligent applications in CS studies.

Conclusion

We have proposed the convenient non-harm CS intelligent identify method EasiCNCSII which includes data acquisition method and intelligent CS identification algorithms. It is able to collect data in a harmless and convenient method in a short period of time and identity the CS with an encouraging performance. To achieve the best performance on the limited data set of weak availability, we developed a data processing framework consisting of feature extraction, feature selection after the extraction, and the intelligent recognition algorithm. Combined with mobile application, the EasiCNCSII potentially provide low-cost convenient universal access to indispensable care outside the hospital, especially in the remote rural areas with poor medical resources, which potentially promotes a balanced distribution of quality medical resources. Our future work will focus on collecting and building the more larger database including sEMG signal data and other CS-related information, and implement an overall high-performance CS identification application system.

Data Availability

The raw sEMG data supporting this study are not publicly available due to user privacy, but are available from the corresponding author on reasonable request. However, the data set, which are extracted from raw sEMG data by the feature extraction methods in this paper, support automatic CS identification study, are available at github after reorganized, each sample of which consist of 2949 features. Currently, it is available from wangnana@ict.ac.cn on reasonable request.

References

1. Côté, P., Cassidy, J. D. & Carroll, L. The saskatchewan health and back pain survey: the prevalence of neck pain and related disability in saskatchewan adults. *Spine* **23**, 1689–1698 (1998).
2. Todd, A. G. Cervical spine: degenerative conditions. *Current reviews in musculoskeletal medicine* **4**, 168 (2011).
3. Jebri, B. *et al.* Detection of degenerative change in lateral projection cervical spine x-ray images. In *Medical Imaging 2015: Computer-Aided Diagnosis*, vol. 9414, 941404 (International Society for Optics and Photonics, 2015).
4. Singh, S., Kumar, D. & Kumar, S. Risk factors in cervical spondylosis. *Journal of Clinical Orthopaedics and Trauma* **5**, 221–226 (2014).
5. Hoggjohnson, S., Velde, G. V. D., Carroll, L. J., Holm, L. W. & Cassidy, J. D. The burden and determinants of neck pain in the general population: Results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders. best evidence on the burden and determinants of neck. *European Spine Journal* **17**, 39–51 (2008).
6. Damian Hoy, L. M. *et al.* The global burden of neck pain: estimates from the global burden of disease 2010 study. *Annals of the rheumatic diseases* **73**, 1309–1315 (2014).
7. Matz, P. G. *et al.* Joint section on disorders of the spine and peripheral nerves of the american association of neurological surgeons and congress of neurological surgeons. *J Neurosurg Spine* **11**, 157–169 (2009).
8. Kotil, K. & Bilge, T. Prospective study of anterior cervical microforaminotomy for cervical radiculopathy. *Journal of Clinical Neuroscience* **15**, 749–756 (2008).
9. Cai, Z. *et al.* Trend of the incidence of cervical spondylosis: decrease with aging in the elderly and increase with aging in the young and the adults. *Int J Clin Exp Med* **9**, 14329–14336 (2016).
10. Quinn, J. C., Kiely, P. D., Lebl, D. R. & Hughes, A. P. Anterior surgical treatment of cervical spondylotic myelopathy. *HSS Journal* **11**, 15–25 (2015).
11. Xie, X., Yin, S., Gao, H. & Kaynak, O. Asymptotic stability and stabilisation of uncertain delta operator systems with time-varying delays. *Control Theory and Applications* **7**, 1071–1078 (2013).
12. Shen, Y., Ding, S. X., Haghani, A., Hao, H. & Zhang, P. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of Process Control* **22**, 1567–1581 (2012).
13. Liu, M., Cui, L., Guan, Y., Li, B. & Du, H. Single-fiber electromyography in amyotrophic lateral sclerosis and cervical spondylosis. *Muscle and nerve* **48**, 137–139 (2013).
14. Borden, S. Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. *J Bone Joint Surg.* **72**, 1178–1184 (1990).
15. Lehto, I. *et al.* Age-related MRI changes at 0.1 T in cervical discs in asymptomatic subjects. *Neuroradiology* **36**, 49–53 (1994).

16. Yu, X., Liu, M., Meng, L. & Xiang, L. Classifying cervical spondylosis based on x-ray quantitative diagnosis. *Neurocomputing* **165**, 222–227 (2015).
17. Binder, A. I. Cervical spondylosis and neck pain. *BMJ: British Medical Journal* **334**, 527 (2007).
18. Chitte, P. P. & Gokhale, U. M. Analysis of different methods for identification and classification of cervical spondylosis (cs): A survey. *International Journal of Applied Engineering Research* **12**, 11727–11737 (2017).
19. Zhongmin, Z. *Study on the Symptoms and Soft Tissue Changes of Cervical spondylosis*. Ph.D. thesis, China Academy of Chinese Medical Sciences (2011).
20. Panjabi, M. M. The stabilizing system of the spine. part i. function, dysfunction, adaptation, and enhancement. *Clinical Spine Surgery* **5**, 383–389 (1992).
21. Airaksinen, M. K. *et al.* Wireless on-line electromyography in recording neck muscle function: a pilot study. *Pathophysiology* **12**, 303–306 (2005).
22. Falla, D. Unravelling the complexity of muscle impairment in chronic neck pain. *Manual therapy* **9**, 125–133 (2004).
23. Falla, D., Bilenkij, G. & Jull, G. Patients with chronic neck pain demonstrate altered patterns of muscle activation during performance of a functional upper limb task. *Spine* **29**, 1436–1440 (2004).
24. Johnston, V., Jull, G., Souvlis, T. & Jimmieson, N. L. Neck movement and muscle activity characteristics in female office workers with neck pain. *Spine* **33**, 555–563 (2008).
25. Madeleine, P., Xie, Y., Szeto, G. P. & Samani, A. Effects of chronic neck–shoulder pain on normalized mutual information analysis of surface electromyography during functional tasks. *Clinical Neurophysiology* **127**, 3110–3117 (2016).
26. Johnston, V., Jull, G., Darnell, R., Jimmieson, N. & Souvlis, T. Alterations in cervical muscle activity in functional and stressful tasks in female office workers with neck pain. *European journal of applied physiology* **103**, 253–264 (2008).
27. Falla, D., Farina, D., Dahl, M. K. & Graven-Nielsen, T. Muscle pain induces task-dependent changes in cervical agonist/antagonist activity. *Journal of Applied Physiology* **102**, 601–609 (2007).
28. Yang, X. *Clinical observation of common tender points in cervical spondylosis*. Ph.D. thesis, Beijing University of Chinese Medicine (2010).
29. Xiaoqian, Y. *Clinical observation of common tender point in cervical spondylosis*. Master's thesis, Beijing University of Chinese Medicine (2010).
30. Association of rehabilitation medicine, C. Guide to diagnosis and treatment of cervical spondylosis. Tech. Rep., Chinese association of rehabilitation medicine (2010).
31. Yang, L.-F. & Kang, B. Study on human neck muscles' comfort of different height levels based on semg method. In *Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation*, 563–574 (Springer, 2016).
32. Venugopal, G., Navaneethakrishna, M. & Ramakrishnan, S. Extraction and analysis of multiple time window features associated with muscle fatigue conditions using semg signals. *Expert Systems with Applications* **41**, 2652–2659 (2014).
33. Yang, H.-S. *et al.* Influence of different acupoint combinations on immediate effect of surface electromyography of patients with cervical spondylosis. *World Journal of Acupuncture-Moxibustion* **26**, 7–13 (2016).
34. Lambelet, C., Lyu, M., Woolley, D., Gasser, R. & Wenderoth, N. The wrist—a wearable wrist exoskeleton with semg-based force control for stroke rehabilitation. In *Rehabilitation Robotics (ICORR), 2017 International Conference on*, 726–733 (IEEE, 2017).
35. Butte, A. J. Big data opens a window onto wellness. *Nature biotechnology* **35**, 720 (2017).
36. de Vries, E. N. *et al.* Effect of a comprehensive surgical safety system on patient outcomes. *New England Journal of Medicine* **363**, 1928–1937 (2010).
37. Reed, M. *et al.* Implementation of an outpatient electronic health record and emergency department visits, hospitalizations, and office visits among patients with diabetes. *Jama* **310**, 1060–1065 (2013).
38. Jaffe, M. G., Lee, G. A., Young, J. D., Sidney, S. & Go, A. S. Improved blood pressure control associated with a large-scale hypertension program. *Jama* **310**, 699–705 (2013).
39. Yoo, C., Ramirez, L. & Liuzzi, J. Big data analysis using modern statistical and machine learning methods in medicine. *International neurology journal* **18**, 50 (2014).
40. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115 (2017).
41. Lezcano-Valverde, J. M. *et al.* Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. *Scientific reports* **7**, 10189 (2017).
42. Zhou, X. X. *et al.* pdeep: Predicting MS/MS spectra of peptides with deep learning. *Analytical Chemistry* (2017).
43. Subasi, A., Alickovic, E. & Kevric, J. Diagnosis of chronic kidney disease by using random forest. In *CMBBIH 2017*, 589–594 (Springer, 2017).
44. Dehmer, M., Emmert-Streib, F., Graber, A. & Salvador, A. *Applied statistics for network biology: methods in systems biology* (John Wiley & Sons, 2011).
45. Dehmer, M. & Emmert-Streib, F. *Statistical Diagnostics for Cancer: Analyzing High-Dimensional Data* (John Wiley & Sons, 2013).
46. Li, J. *et al.* Feature selection: A data perspective. *ACM Computing Surveys (CSUR)* **50**, 94 (2017).
47. Zhou, Z.-H. Ensemble learning. *Encyclopedia of biometrics* 411–416 (2015).
48. Li, P. Robust logitboost and adaptive base class (abc) logitboost. *arXiv preprint arXiv:1203.3491* (2012).
49. Burges, C. J. From ranknet to lambdarank to lambdamart: An overview. *Learning* **11**, 81 (2010).
50. Bennett, J. *et al.* The netflix prize. In *Proceedings of KDD cup and workshop*, vol. 2007, 35 (New York, NY, USA, 2007).
51. He, X. *et al.* Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, 1–9 (ACM, 2014).
52. Altun, K., Barshan, B. & Tunçel, O. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition* **43**, 3605–3620 (2010).
53. Yükksek, M. C. A comparative study on human activity classification with miniature inertial and magnetic sensors. *Ph.D. thesis, bilkent university* (2011).
54. Chowdhury, S. K., Nimbarte, A. D., Jaridi, M. & Creese, R. C. Discrete wavelet transform analysis of surface electromyography for the fatigue assessment of neck and shoulder muscles. *Journal of Electromyography and Kinesiology* **23**, 995–1003 (2013).
55. Babita, Kumari, P., Narayan, Y. & Mathew, L. Binary movement classification of semg signal using linear svm and wavelet packet transform. In *IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems*, 1–4 (2017).
56. Cui, J. G., Wang, X., Zhong hai, L. I. & Tian, F. The method of surface emg pattern recognition based on ar parameter model and clustering analysis. *Acta Metrologica Sinica* **27**, 286–289 (2006).
57. Chen, W. T., Wang, Z. Z., Xiao, H. U. & Xiao-Pu, L. I. Entropy analysis of semg signal during dynamic contractions for assessing muscle fatigue. *Chinese Journal of Medical Physics* (2006).
58. Shin, S.-J., An, D.-H., Oh, J.-S. & Yoo, W.-G. Changes in pressure pain in the upper trapezius muscle, cervical range of motion, and the cervical flexion–relaxation ratio after overhead work. *Industrial health* **50**, 509–515 (2012).
59. Murphy, B. A., Marshall, P. W. & Taylor, H. H. The cervical flexion-relaxation ratio: reproducibility and comparison between chronic neck pain patients and controls. *Spine* **35**, 2103–2108 (2010).
60. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794 (ACM, 2016).
61. Baldi, P., Sadowski, P. & Whiteson, D. Searching for exotic particles in high-energy physics with deep learning. *Nature communications* **5**, 4308 (2014).

62. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern classification* (John Wiley & Sons, 2012).
63. Brown, G., Pocock, A., Zhao, M.-J. & Luján, M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *Journal of machine learning research* **13**, 27–66 (2012).
64. Cai, D., Zhang, C. & He, X. Unsupervised feature selection for multi-cluster data. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 333–342 (ACM, 2010).
65. Piwek, L., Ellis, D. A., Andrews, S. & Joinson, A. The rise of consumer health wearables: promises and barriers. *PLoS Medicine* **13**, e1001953 (2016).

Acknowledgements

We thank Dr. Yunyou Huang for his help with the experimental design, data acquisition, data processing, and Fanda Fan for the help with the GBRT. We also thank Dr. Bingyan Cao and the other medical workers in Beijing Xiyuan Hospital rehabilitation physiotherapy for recruiting the subjects (the outpatients and the healthy free from the CS) and Research Center for Ubiquitous Computing Systems (in short CUbiCS) for the recruiting the subjects (the healthy). Besides, the authors also thank all subjects for providing valuable data. The paper was supported by the National Natural Science Foundation of China (NSFC) under Grant No. 61672498. The study was supported by the National Natural Science Foundation of China (NSFC) under Grant No. 61672498.

Author Contributions

N.-N.W. designed the experiment and model, collected the data, wrote the codes, analyzed the data and wrote the manuscript. L.C. directed the project and revised the manuscript. X.H. guided the data preparation. Y.R. and J.X. participate in recruiting subjects and provide places for data acquisition. J.H.L. participated in the data acquisition. N.W. provide the laptop and sEMG device.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-32377-3>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018