Final report: Automation of the DXA Scoliosis Method

PI: Dr Emma Clark, University of Bristol

Collaborators: From University of Oxford - Dr Amir Jamaludin, Dr Timor Kadir, Prof Andrew Zisserman, Prof Jeremy Fairbank. From North Bristol NHS Trust – Mr Ian Harding.

Duration of funding: January 2018 for 12 months

…….

Executive Summary

This project aimed to develop and validate a fully automatic system to identify and measure spinal curvature from total body DXA scans.

The data was split into a training, validation and test sets. The new software algorithm was developed from modification of an established system. Training involved prediction of segmentation masks, production of mid-spine maps, and final classification into scoliosis or no scoliosis based on suspiciousness scores. The final model was interrogated using heat maps to identify which area of the total body DXA scans was contributing to the classification.

The final automated model has an accuracy of 94.2%, a sensitivity of 86.0%, a specificity of 95.8% and an area under the receiver operator curve value (AUC) of 0.91 (95%CI 0.87 to 0.95). Disagreement between the automated model and the manual annotation is likely to be explained by errors with the original manual annotation in at least half the cases.

Next steps include further model development to allow quantification of curve size, clinical validation based on ALSPAC, assessment of the natural history of AIS at a population-level, development of a clinical tool to predict which people with AIS progress, and similar work on adult degenerative scoliosis using UK BioBank.

One paper has been published so far, and two more are planned. A collaboration has been set up with Augment (a project using UK BioBank) to use the automated model on 100,000 adult DXA scans. Further grants are planned.

BACKGROUND

Spinal curvature (scoliosis) affects 3-6%[1] of the population. Scoliosis generally appears in children while they are growing. The curves can get worse both while growing and more slowly with aging. New scoliosis also occurs in adults (called degenerative scoliosis). Large curves are associated with significant problems[2-4] and may require surgery[5]. Little is known about the causes of curve onset, progression and prognosis. What is needed is a clinical prediction tool to decide which people with scoliosis are most likely to progress and need ongoing monitoring, and who can be reassured and discharged. However, development of a prediction tool needs large scale population-based research to identify predictors. This has been limited so far, because the traditional method of identifying scoliosis is to carry out spinal X-rays which impart a high dose of radiation[6], and are not ethical for screening the whole population.

To address this, we have developed and validated a method of measuring scoliosis from standard total body dual energy X-ray absorptiometry (DXA) scans[7], a low-radiation technique using data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a population-based birth

cohort. We have then applied our DXA Scoliosis Method (DSM) to 7000 DXA scans from the ALSPAC aged 9 research clinic and 5000 DXA scans from the aged 15 research clinic, and have identified novel associations between early life physical activity[8] and body composition[9] with onset of scoliosis by aged 15. However, the prevalence of scoliosis at aged 15 is relatively low (7.9% in ALSPAC) and to carry out appropriately powered epidemiological studies we need to combine data from multiple research cohorts. We have identified additional research cohorts that have total body DXA scans already performed (approximately 84,500 individuals). We plan to use these cohorts to generate a prediction tool for scoliosis onset and progression. However, our manual method for identifying and measuring scoliosis from DXA scans is labour intensive, and further research is not feasible until we automate this method.

Aim of this grant

Our grant therefore proposed development and validation of a fully automatic system to identify and measure spinal curvature from total body DXA scans based on our manual method. The engineering department at the University of Oxford has already developed a similar automated system for three-dimension images of the spine based on magnetic resonance imaging (MRI) scans. We planned to modify this system to allow automatic collection of data on spinal scoliosis from total body DXA scans for future research purposes. This modification process was planned to have two stages: (1) development of a new software algorithm to extract the required features and classify spinal images based on a subset of anonymised DXA images from ALSPAC; and (2) validation of the software on a further dataset of anonymised images from ALSPAC.

METHODS

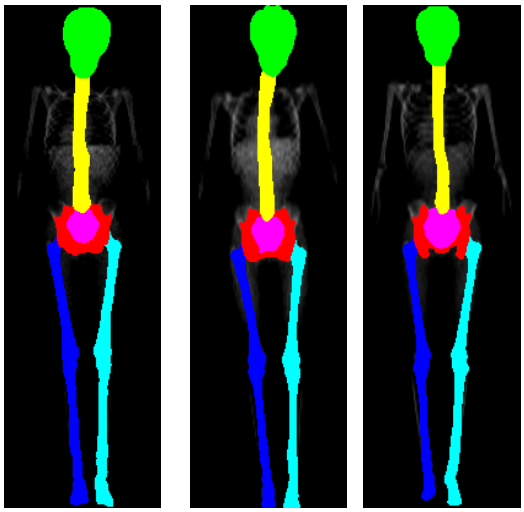
This project was co-ordinated by Dr Emma Clark at Bristol, but the automation was developed and tested by the Department of Engineering at the University of Oxford. The research group met every 2-3 months face-to-face.

The ALSPAC DXA data at aged 9 and 15 was split into a training, a validation and a test set.

Stage 1: Development of the new software algorithm

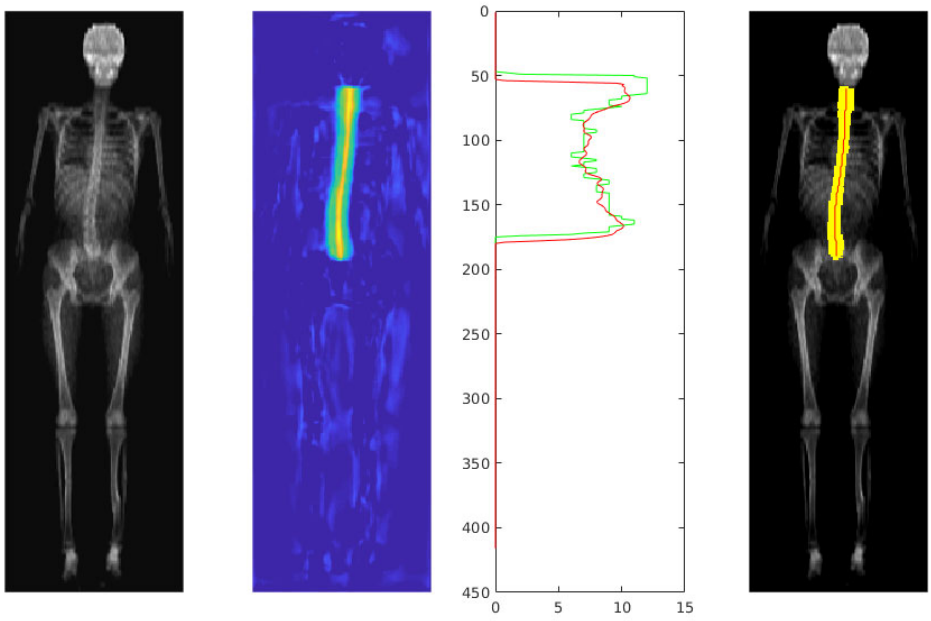
Using the training set, all images were standardised to the same height keeping the aspect ratio (isotropic scaling using the SpineNet software). Strong labels were developed to help the model understand spinal anatomy (see Figure 1).

Figure 1: Prediction of the strong labels (segmentation masks)



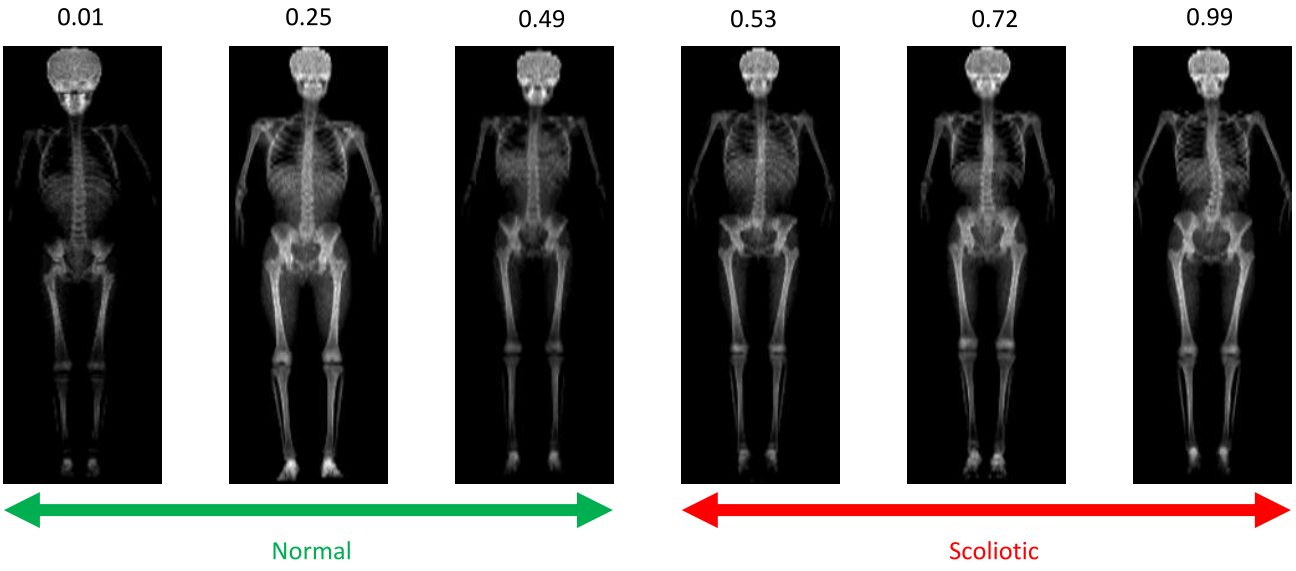
The model then produced mid-spine maps (see Figure 2)

Figure 2: An example of the mid-spine map produced by the model



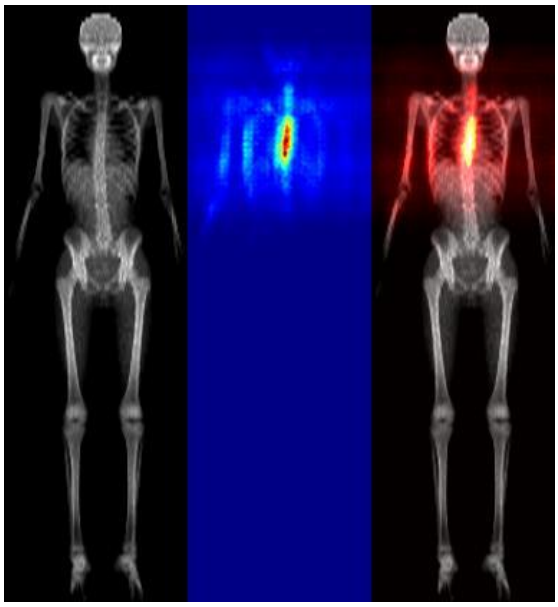
The model was then trained against the manual classifier of scoliosis/no scoliosis based on the previously validated manual cut-off[7]. Each classification produced by the model comes with a score – the so-called 'suspiciousness score'. This score ranges from 0 (normal) to 1 (scoliosis) – see Figure 3

Figure 3: Scoliosis scores produced by the automated model



Accuracy of the model was then improved through modifications to labels, maps, and by summing the scoliotic classes post-softmax. To interrogate the model to identify 'how' it was making the decisions heatmaps were produced (see Figure 4).

Figure 4: Heatmaps produced by the model indicating the site of the total body DXA scans that contributed to the decision that scoliosis was present



Stage 2: Validation of the model

Using the validation set, all images were standardised to the same height keeping the aspect ratio (isotropic scaling using the SpineNet software). The automation was run and accuracy of the output was calculated by averaging the recall obtained for the two different classes; scoliosis/no scoliosis.

Using the test set, all images were standardised to the same height keeping the aspect ratio (isotropic scaling using the SpineNet software). The automation was run and the validity of the output was calculated by comparison with the manual DSM annotation using sensitivity, specificity and area under the receiver operator curve values (AUC). Test time augmentation was used for the prediction whereby for each scan, the average prediction is obtained by averaging the scores of the original score and its horizontal flip.

Where discrepancies between the automated model and manual annotations occurred, a random selection of images were reviewed by clinicians (Jeremy Fairbank, Ian Harding and Emma Clark) to identify if the manual annotations were correct or incorrect. In addition, the automated model was run on the aged 17 scans for the discrepancies to identify if the images were still classified as scoliosis by the model, and these aged 17 scans were also reviewed by the clinicians.

RESULTS

The accuracy of the automated model at identifying scoliosis in the test set is 94.2% ± 2.1. Using a cut-off of suspiciousness score of 0.95, compared to the manual DSM, the automated model has a sensitivity of 86.0%, a specificity of 95.8% and an area under the receiver operator curve value (AUC) of 0.91 (95%CI 0.87 to 0.95) – see Table 1.

Table 1: Automated prediction of scoliosis >6 degrees with suspiciousness score of 0.95 and above, compared to manual prediction of scoliosis (DSM) >6 degrees

	Manual method		
Automation	No scoliosis	Scoliosis	
No scoliosis	638	13	651
Scoliosis	28	80	108
	666	93	759

Assessment of discrepancies: no scoliosis by manual method, scoliosis by automated model

A random sample of the 28 scans where the manual method did not identify scoliosis but the automated model did, were reviewed by three clinicians (two spinal surgeons and one rheumatologist who developed the manual method). All were unaware of the automated prediction, and all read the scans without knowledge of each others interpretation.

Of the 9 scans reviewed, 5 were identified as having scoliosis (in agreement with the automated model) by all three clinicians, suggesting the manual annotation was incorrect in these cases. There was not agreement for the other 4 scans as to whether scoliosis was present or not due to small size of spinal abnormality and body positioning.

Assessment of discrepancies: scoliosis by manual method, no scoliosis by automated model

A random sample of the 13 scans where the manual method did identify scoliosis but the automated model did not, were reviewed by three clinicians (two spinal surgeons and one rheumatologist who developed the manual method). All were unaware of the automated prediction, and all read the scans without knowledge of each others interpretation.

Of the 10 scans reviewed, 6 were identified as not having scoliosis (in agreement with the automated model) by all three clinicians, suggesting the manual annotation was incorrect in these cases. There was not agreement for the other 4 scans as to whether scoliosis was present or not due to small size of spinal abnormality and body positioning.

Assessment discrepancies: comparison with automated model prediction on aged 17 data

The automated model was run on the aged 17 images for those randomly selected discrepancy scans described above. 17 out of 19 participants had scans at aged 17. In 14/17 (82%) of participants, the automated model classified their spines the same at aged 15 and aged 17, thereby increasing the confidence that the model output is valid.

Conclusion

The automated model has good accuracy, sensitivity, specificity and AUC. Disagreement between the automated model and the manual annotation is likely to be explained by errors with the original manual annotation in at least half the cases.

NEXT STEPS

Clinical validation

Clinical validation is required, and this will be carried out over the next few months using ALSPAC. The final output of the automated model will be combined with the rest of the ALSPAC data, and epidemiological analyses will be undertaken to identify if similar associations already found using the manual DSM are also found using the automated model. This will include prevalence and gender distribution[7], the association with altered body composition[9], the association with reduced physical activity/ability[8], and the impact in terms of future back pain and time off school[10]. This, combined with the validation data presented above, will be published in a peer-reviewed journal, and we are aiming for Calcified Tissue International.

Work on quantification of curve size

The model will be developed to quantify curve size with an aim of identifying if one curve is bigger than another.

Assessment of natural history of AIS at a population-level

The automated model will be run on ALSPAC participants at aged 11, 13 and 17 years, and a descriptive analysis will be undertaken to describe the natural history of AIS in terms of progression, regression and stabilisation. Trajectories will be plotted with confidence intervals. We could then define different populations across ages 9 to 17 (for example no progression, trivial progression, serious progression) which would have potential practical applications. This will be published in a peer-reviewed journal.

Development of a clinical tool to predict which people with AIS progress

Further funding will be sought from the Medical Research Council (MRC) to run the automated model on other population-based birth cohorts around the world with repeated total body DXA measures already collected. This will give us enough power to develop a predictive model to identify which people progress and which stabilise.

Similar work on adult degenerative scoliosis

We are working on a collaboration with Augment led by Prof Jon Tobias at the University of Bristol to run the automated model on the total body DXA scans being done by UK BioBank (n=100,000). This will allow us to carry out one of the first population-based analyses of adult degenerative scoliosis.

OUTPUTS OF THIS GRANT

1. Jamaludin A, Kadir T, Clark EM, Zisserman A. Predicting scoliosis in DXA scans using intermediate representation (2018) Computational Methods and Clinical Applications for Spine Imaging, MICCAI Workshop – to be published as post-proceedings in Springer's Lecture Notes in Computer Science (See Appendix)

2. clinical validation paper – aim for CTI, using data from aged 9 and 15 (same as original manual paper) plus possible data from aged 11 or 13 – in preparation

3. natural history/descriptive paper using data from aged 9,11,13,15 and 17 – how many people develop scoliosis, progress or regress. Look at known predictors already identified from ALSPAC: body composition, adiponectin, physical activity +/- angles and some basic descriptors – to be prepared

REFERENCES

1. Grivas, T.B., et al., *School screening in the heavily industrialised area - is there any role of industrial environmental factors in idiopathic scoliosis prevalence?* Studies in Health Technology and Information, 2002. **91**: p. 76-80.
2. Miller, N.H., *Cause and natural history of AIS.* Orthopedic Clinic of North America, 1999. **30**(3): p. 343-352.
3. Aaro, S. and C. Ohlund, *Scoliosis and pulmonary function.* Spine, 1984. **9**: p. 220.
4. Weinstein, S.L. and I.V. Ponseti, *Idiopathic scoliosis: Long term follow-up and prognosis in untreated patients.* Journal of Bone and Joint Surgery - American Volume, 1981. **63**: p. 702-711.
5. Weinstein, S.L., et al., *Adolescent Idiopathic Scoliosis.* Lancet, 2008. **371**: p. 1527-1537.
6. Watson, S.J., et al., *Ionising radiation exposure of the UK population.* 2005 Review, 2005. **HPA-RPD-001**.
7. Taylor, H.J., et al., *Identifying scoliosis in population-based cohorts: Development and validation of a novel method based on total body DXA scans.* Calcified Tissue International, 2013. **92**(6): p. 539-547.
8. Tobias, J.H., et al., *Association between physical activity and scoliosis: a prospective cohort study.* International Journal of Epidemiology, 2018: p. doi.org/10.1093/ije/dyy268.
9. Clark, E.M., et al., *Association between components of body composition and scoliosis: A prospective cohort study reporting differences identifiable before the onset of scoliosis.* Journal of Bone and Mineral Research, 2014. **29**(8): p. 1729-1736.
10. Clark, E.M., J.H. Tobias, and J. Fairbank, *The impact of small spinal curves in adolescents that have not presented to secondary care: a population-based cohort study.* Spine, 2016. **41**(10): p. E611-E617.

Appendix:

Predicting Scoliosis in DXA Scans Using Intermediate Representations

Amir Jamaludin[1], Timor Kadir[2], Emma Clark[3], and Andrew Zisserman[1]

[1] VGG, Dept. of Engineering Science, University of Oxford
[2] Optellum
[3] Musculoskeletal Research Unit, School of Clinical Sciences, University of Bristol
{amirj,az}@robots.ox.ac.uk
timor.kadir@optellum.com
emma.clark@bristol.ac.uk

Abstract. We describe a method to automatically predict scoliosis in Dual-energy X-ray Absorptiometry (DXA) scans. We also show that intermediate representations, which in our case are segments of body parts, help improve performance. Hence, we propose a two step process for prediction: (i) we learn to segment body parts via a segmentation Convolutional Neural Network (CNN), which we show outperforms the noisy labels it was trained on, and (ii) we predict with a classification CNN that uses as input both the raw DXA scan and also the intermediate representation, i.e. the segmented body parts. We demonstrate that this two step process can predict scoliosis with high accuracy, and can also localize the spinal curves (i.e. geometry) without additional supervision. Furthermore, we also propose a soft score of scoliosis based on the classification CNN which correlates to the severity of scoliosis.

1 Introduction

Scoliosis is an abnormal sideways curvature of the spine typically occuring prior to puberty and affects approximately 3% of the population. While most cases are mild, stablizing over time and presenting few symptoms, some children develop severe deformaties that can cause lifelong disability and pain. Scoliosis can also cause back pain [1] and in rare cases can cause respiratory failure [8]. It is not currently possible to determine prognosis at the onset of disease and hence children with scoliosis are monitored with repeated X-Ray imaging to determine whether the disease is stable or progressing. While accepted as the standard of care, the use of repeated X-Ray imaging on children with the associated radiation dose is far from ideal. Moreover, the radiation dose also precludes its use in population based epidemiological studies to better understand disease progression and develop future tools to predict prognosis and for screening.

DXA Scans: The use of DXA imaging for diagnosis and monitoring of scoliosis has been proposed as an alternative to X-Ray due to its very low radiation dose compared to spinal X-Rays (0.001 mSv vs. 1.5 mSv) and widespread availability [12]. DXA scans, typically used to measure bone mineral density in the management of osteoporosis, are whole body scans acquired in a line scanning manner from the top of the head to the bottom of the feet. Two X-Ray sources at different energy levels are used to create a pair of absorption images which are then post-processed to produce quantitative bone mineral density images. While detection of scoliosis using DXA has been shown to be feasible and accurate, the manual technique proposed by [12] is labour intensive and requires