HHS Public Access

Author manuscript

*Proc SPIE Int Soc Opt Eng*. Author manuscript; available in PMC 2019 November 18.

Published in final edited form as:

*Proc SPIE Int Soc Opt Eng*. 2018 February ; 10575: . doi:10.1117/12.2293023.

# Quantifying the Association between White Matter Integrity Changes and Subconcussive Head Impact Exposure from a Single Season of Youth and High School Football using 3D Convolutional Neural Networks

**Behrouz Saghafi**[a], **Gowtham Murugesan**[a], **Elizabeth Davenport**[a], **Ben Wagner**[a], **Jillian Urban**[b], **Mireille Kelley**[b], **Derek Jones**[b], **Alexander Powers**[b], **Christopher Whitlow**[b], **Joel Stitzel**[b], **Joseph Maldjian**[a], **Albert Montillo**[a]

[a]University of Texas Southwestern Medical Center, Dallas, TX, USA

[b]Wake Forest School of Medicine, Winston-Salem, NC, USA

## Abstract

The effect of subconcussive head impact exposure during contact sports, including American football, on brain health is poorly understood particularly in young and adolescent players, who may be more vulnerable to brain injury during periods of rapid brain maturation. This study aims to quantify the association between cumulative effects of head impact exposure from a single season of football on white matter (WM) integrity as measured with diffusion MRI. The study targets football players aged 9–18 years old. All players were imaged pre- and post-season with structural MRI and diffusion tensor MRI (DTI). Fractional Anisotropy (FA) maps, shown to be closely correlated with WM integrity, were computed for each subject, co-registered and subtracted to compute the change in FA per subject. Biomechanical metrics were collected at every practice and game using helmet mounted accelerometers. Each head impact was converted into a risk of concussion, and the risk of concussion-weighted cumulative exposure (RWE) was computed for each player for the season. Athletes with high and low RWE were selected for a two-category classification task. This task was addressed by developing a 3D Convolutional Neural Network (CNN) to automatically classify players into high and low impact exposure groups from the change in FA maps. Using the proposed model, high classification performance, including ROC Area Under Curve score of 85.71% and F1 score of 83.33% was achieved. This work adds to the growing body of evidence for the presence of detectable neuroimaging brain changes in white matter integrity from a single season of contact sports play, even in the absence of a clinically diagnosed concussion.

## 1. INTRODUCTION

Organized Football is played by 5 million athletes in the United States, with 96% of those players at the youth and high school level.[1] Its popularity is underscored by the National Federation of State High School Associations (NFHS), who found football to be one of the most frequently played sport among high school students.[2] Despite the large number of

adolescent players, most research on the effect of football head impacts have focused on the collegiate and professional levels. Given that the adolescent brain is still undergoing rapid development this omission could have significant consequences. Additionally, the majority of football studies have focused on changes induced from concussions while relatively little attention has been paid to study the impact of repetitive subconcussive impacts, though these are much more common in the sport. This study aims to alleviate these gaps by studying the effects of repetitive subconcussive head impact exposure in the youth and high school level. Specifically the goal is to quantify the association between changes in white matter integrity and exposure to cumulative subconcussive head impacts during a single season of football.

## 2. MATERIALS AND METHODS

This work is part of an ongoing IRB-approved study called Imaging Telemetry And Kinematic modeLing in football (iTAKL) [3, 4]. The players range in age from 9 to 18 years old. Each player was instrumented with the Head Impact Telemetry System (HITS)[3] in which accelerometers are mounted inside each players helmet to measure linear and rotational accelerations. The accelerometers are spring mounted, keeping them in constant contact with the skull during head impacts. The biomechanical metrics collected from HITS over the season were used to compute the risk of concussion-weighted cumulative exposure (RWE)[4] from both linear and rotational head impact accelerations.

Diffusion-weighted as well as T1-weighted structural MRI scans were collected from all players pre- and post-season using a 3.0 Tesla Siemens Skyra MRI scanner. Diffusion-weighted images were acquired with spatial resolution of $2.2\times2.2\times2.2$ $mm^3$. Diffusion-weighting (DW) was applied along 15 directions with $b = 1000$ $s/mm^2$. In addition, ten B0 images were obtained and the DW images were eddy current corrected by normalizing each image to the B0 image via mutual information using the FSL package.[5] Fractional Anisotropy (FA) maps were computed for each image using the DTI-TK library.[6] The B0 images were co-registered to T1-weighted images using an affine transformation. Then T1-weighted images were normalized to MNI space using the DARTEL non-linear transform[7] computed via the VBM8 toolbox*. These transforms were used to spatially normalize the FA maps into MNI space.

The imaging data were quality checked and any data with prominent imaging artifact or severe head motion was not used. Subjects with a history of concussion or clinically significant imaging findings were excluded. This yielded 122 players.

ΔFA volumetric maps were computed by subtracting the pre-season FA map from the post-season FA map. The ΔFA maps were cropped to encompass the brain volume only and downsampled to resolution of $3 \times 3 \times 3$ $mm^3$ using cubic interpolation. This yielded image volumes with dimensions of $48 \times 60 \times 52$ voxels.

Each player's level of head impact exposure for the season was summarized with a single measure known as the cumulative Risk of concussion-Weighted Exposure (RWE) score, by

*http://dbm.neuro.uni-jena.de/vbm.html

converting the accelerometer measures for each hit into a risk of concussion and summing these impact risks for the season.[4] Players were then dichotomized based on this seasonal head impact exposure level, RWE. The 24 players with RWE>1.11 were labeled as high impact exposure players, while the 36 players with RWE<0.13 were labeled as the low impact exposure players. To perform machine learning model selection, 48 of the total 60 players were randomly selected to be used as the aggregate training set while the remaining 12 subjects were held out as the test set and not used during model selection.

The aggregate training set was further divided into a training set and validation set via 5-fold stratified cross-validation. In each fold, ~38 subjects were used for training and ~10 subjects for validation. In order to distinguish the high and low impact exposure players using the ΔFA maps, a 3D Convolutional Neural Network (3D-CNN) classifier was developed. This model automatically learns the best hierarchy of features from 3D image intensities (ΔFA maps) using convolutional layers and then learns to combine them for classification using dense fully-connected layers. Like VGGNet, our network uses small filters of size 3 × 3 × 3 voxels. Innovation occurs through the development and optimization of a problem specific architecture. The number of filters was increased in each successive hidden layer and the number of filters and number of layers were chosen experimentally using the training set (not the test set). Dropout[8] was employed in the dense fully-connected layer to improve the generalization. Compared to other non-linear activation functions (ReLU, hyperbolic tangent, sigmoid), the Parametric Rectified Linear Unit (PReLU)[9] was found to work best. It is applied as the activation function for all the hidden layers. PReLU is defined as, where $\alpha$ is learned from the data. A categorical output layer consisting of a single neuron per category was implemented via the softmax activation function defined as $f(x) = \begin{cases} \alpha x & x < 0 \\ x & x \geq 0 \end{cases}$, where $\alpha$ is learned from the data. A categorical output layer consisting of a single neuron per category was implemented via the softmax activation function defined as $S_j(z) = \frac{e^z j}{\sum_{k=1}^{2} e^z k}$; $j = 1, 2$. Categorical cross-entropy was used as the loss function. Each class was proportionally weighted in the loss function to handle the unbalanced class sample ratios. The ADAM optimization method[10] was used with $\beta_1 = 0.7$, $\beta_2 = 0.999$, and $\epsilon = 1\times10^{-8}$ and weights were initialized to small random values near zero. A batch size of 10 and initial learning rate of $1 \times 10^{-4}$ was chosen based on empirical evidence. An adaptive learning rate was devised. In each validation test, the learning rate was reduced by 50% if the ROC Area Under Curve (AUC) did not improve for 10 epochs. An early stopping schedule with look ahead was developed. Training stops when the network shows no improvement in AUC score for 50 epochs. The winning 3D-CNN architecture that combines all of these strategies is illustrated in figure 1. The model contains 3 stages of convolution, with max pooling used in the end of each stage. Each stage contain 2 convolution layers with the same number of filters. 8 filters is used in the 1st and 2nd layers, 16 filters in the 3rd and 4th layers and 32 filters in the 5th and 6th layers. The convolution stages are followed by 2 dense fully-connected layers with sizes of 32 and 8 respectively and then the output layer computes the probability of the input belonging to each class. Table 1 summarizes the parameters of the proposed architecture.

To perform model selection, the AUC score was computed from 5-fold cross-validation on the training data, and the model that achieved the best (highest) AUC score, 77.50% (±3.33%), was selected. After selecting the proposed architecture, it was trained on the full training set and evaluated on the unseen held-out test set.

## 3. RESULTS

The proposed model, when trained on the full training set (not test) achieves an AUC score of 85.71% and an F1 score of 83.33% with a sensitivity of 100% and specificity of 71.43% on the held-out test set (Table 2). This held-out test set was not used for model selection or model fitting.

The statistical significance of the proposed model was further tested by performing permutation analysis. The null hypothesis is that the CNN model cannot learn to predict the RWE score level based on the training set. The test statistics chosen were the AUC ROC score as well as F1-score on the unseen test set of 12 samples. The permutation testing procedure is as follows:

1. Repeat $R = 400$ times:
    a. Randomly permute the $N$ classification labels over the $N$ subjects.
    b. Compute the value of the test statistic for the current permutation.
2. Construct an empirical probability distribution function (PDF) of the test statistic.
3. Compute the p-value of the test static without permutation.

The PDF for the accuracy test statistics are shown in figure 2. Upon evaluation the proposed model achieved statistically significant reliability; the probability of observing a classifier with higher accuracy than the proposed model is $<1\%$ ($p = 0.0025$). Thus with a significance level of $\alpha = 0.01$, we reject the null hypothesis in favor of the alternative hypothesis that the model has learned to predict the high and low impact exposure categories.

To reveal the regions the proposed classifier found important to distinguish head impact exposure, occlusion maps were computed for the trained CNN.[11] In particular this is achieved by computing, for a given correctly classified subject, the regions important for making that correct classification. This is achieved by computing the decrease in the predicted probability of the correct class when each cuboidal region in the input volume is individually replaced with new random values. A disjoint set of cuboidal regions spanning the entire volume is tested. When a region important for making the prediction is replaced with random intensities, the probability of the high impact exposure prediction will drop substantially. Thus feature importance is directly proportional to this decrease. Such feature importance maps have been created for each 6mm × 6mm × 6mm cuboidal region.

Shown in figure 3, are the occlusion maps for two, correctly classified high impact exposure players with highest predicted class probability. The background shows the FA map for the player, post-season, using a gray scale color map. The overlay shows the importance of the

cuboidal regions using a black to green colormap with brighter green indicating more importance. These are the regions with greater decrease in the probability of the prediction of the correct impact exposure class. In the axial view regions on opposite sizes of the head are selected, while the sagittal view shows the preferences for using the change in deep, centrally located structures. Such structures near the brain's center of gravity, have been shown to receive the highest strain forces in head impact exposure.[12] The patterns of regions used to make the correct high impact exposure prediction are similar but not identical for these players. Further investigation will be useful to reveal the association of these features with the player's specific spatial profile of head impact exposure.

## 4. CONCLUSION

This paper quantifies the association between head impact exposure and changes in diffusion MRI in a single season of football. Since each player was imaged immediately before and after the 3-month season, each player served as its own control and this greatly decreases the influence of potential confounders. While further research with age and BMI matched groups may prove additionally insightful, the results confirm the presence of detectable changes in white matter integrity, as estimated by FA from diffusion MRI, between low and high impact exposure groups of adolescent football players. A 3D convolutional neural network is proposed that classifies players into high and low impact exposure categories. This deep learning approach achieves significant reliability and high classification performance with ROC AUC of 85.71%. This work adds to the growing body of evidence for neuroimaging measurable brain changes to the white matter arise during a single season of contact sports play, even in the absence of a clinically diagnosed concussion.

## ACKNOWLEDGMENTS

This material is based upon work supported by the NSF Graduate Research Fellowship under Grant no. DGE-0907738. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the NSF. Support for this research was also provided by NIH NS082453, NS088125, NS091602 (JM, CW).

## REFERENCES

[1]. Daniel RW, Rowson S, and Duma SM, "Head impact exposure in youth football," Annals of biomedical engineering 40(4), 976–981 (2012). [PubMed: 22350665]

[2]. NFHS, "2015–16 high school athletics participation survey," (2016).

[3]. Crisco JJ, Fiore R, Beckwith JG, Chu JJ, Brolinson PG, Duma S, McAllister TW, Duhaime A-C, and Greenwald RM, "Frequency and location of head impact exposures in individual collegiate football players," Journal of athletic training 45(6), 549–559 (2010). [PubMed: 21062178]

[4]. Urban JE, Davenport EM, Golman AJ, Maldjian JA, Whitlow CT, Powers AK, and Stitzel JD, "Head impact exposure in youth football: high school ages 14 to 18 years and cumulative impact analysis," Annals of biomedical engineering 41(12), 2474–2487 (2013). [PubMed: 23864337]

[5]. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens, et al., "Advances in functional and structural mr image analysis and implementation as fsl," Neuroimage 23, S208–S219 (2004). [PubMed: 15501092]

[6]. Zhang H, Yushkevich PA, Alexander DC, and Gee JC, "Deformable registration of diffusion tensor mr images with explicit orientation optimization," Medical image analysis 10(5), 764–785 (2006). [PubMed: 16899392]

[7]. Ashburner J, “A fast diffeomorphic image registration algorithm,” Neuroimage 38(1), 95–113 (2007). [PubMed: 17761438]

[8]. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, and Salakhutdinov R, “Dropout: a simple way to prevent neural networks from overfitting.,” Journal of Machine Learning Research 15(1), 1929–1958 (2014).

[9]. He K, Zhang X, Ren S, and Sun J, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in [Proceedings of the IEEE international conference on computer vision], 1026–1034 (2015).

[10]. Kingma D and Ba J, “Adam: A method for stochastic optimization,” arXiv preprint arXiv: 1412.6980 (2014).

[11]. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, and Batra D, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” arXiv preprint arXiv: 1610.02391 v3 (2016).

[12]. Bailes JE, Petraglia AL, Omalu BI, Nauman E, and Talavage T, “Role of subconcussion in repetitive mild traumatic brain injury: a review,” Journal of neurosurgery 119(5), 1235–1245 (2013). [PubMed: 23971952]

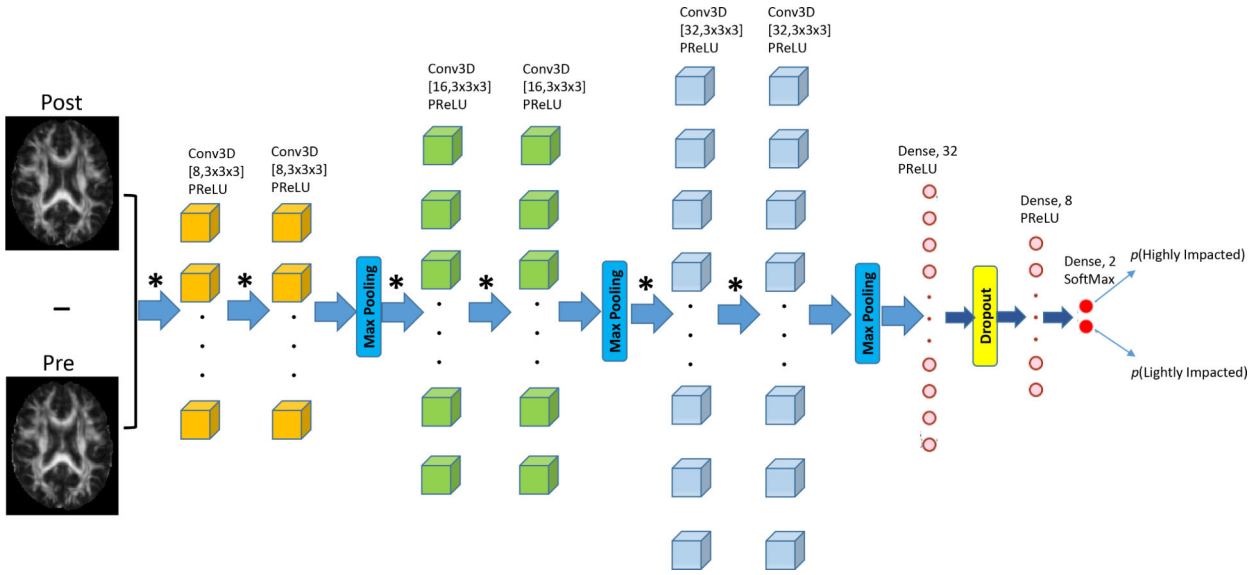


**Figure 1.**

Our proposed deep 3D convolutional neural network for the classification of ΔFA maps into high and light head impact exposures. The cubes represent the 3D convolution filters and the circles represent dense fully-connected neurons. * indicates the convolution operation.

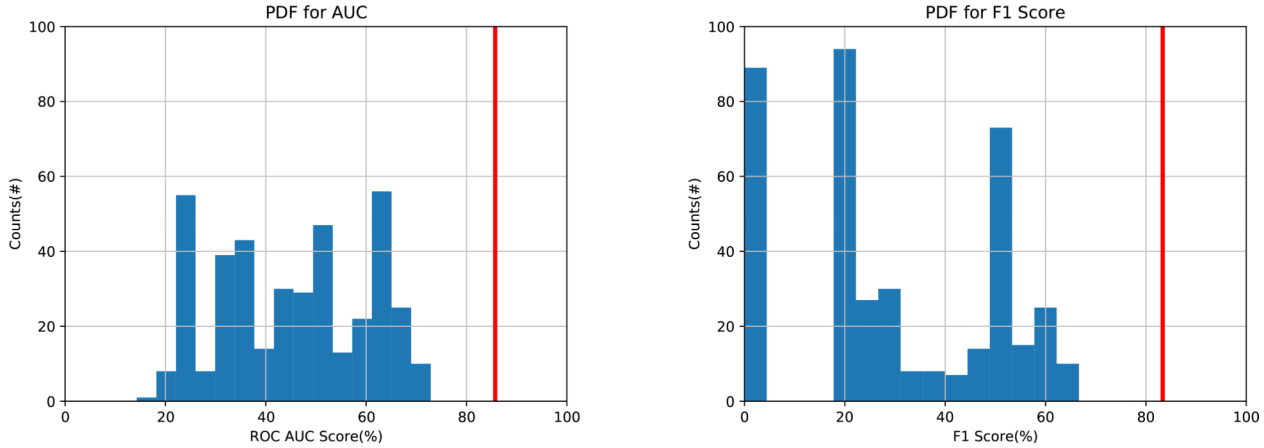


**Figure 2.**

Probability distribution function (PDF) for ROC AUC score and F1-score from permutation analysis for the proposed 3D CNN model. The red lines indicate the scores obtained by the model.

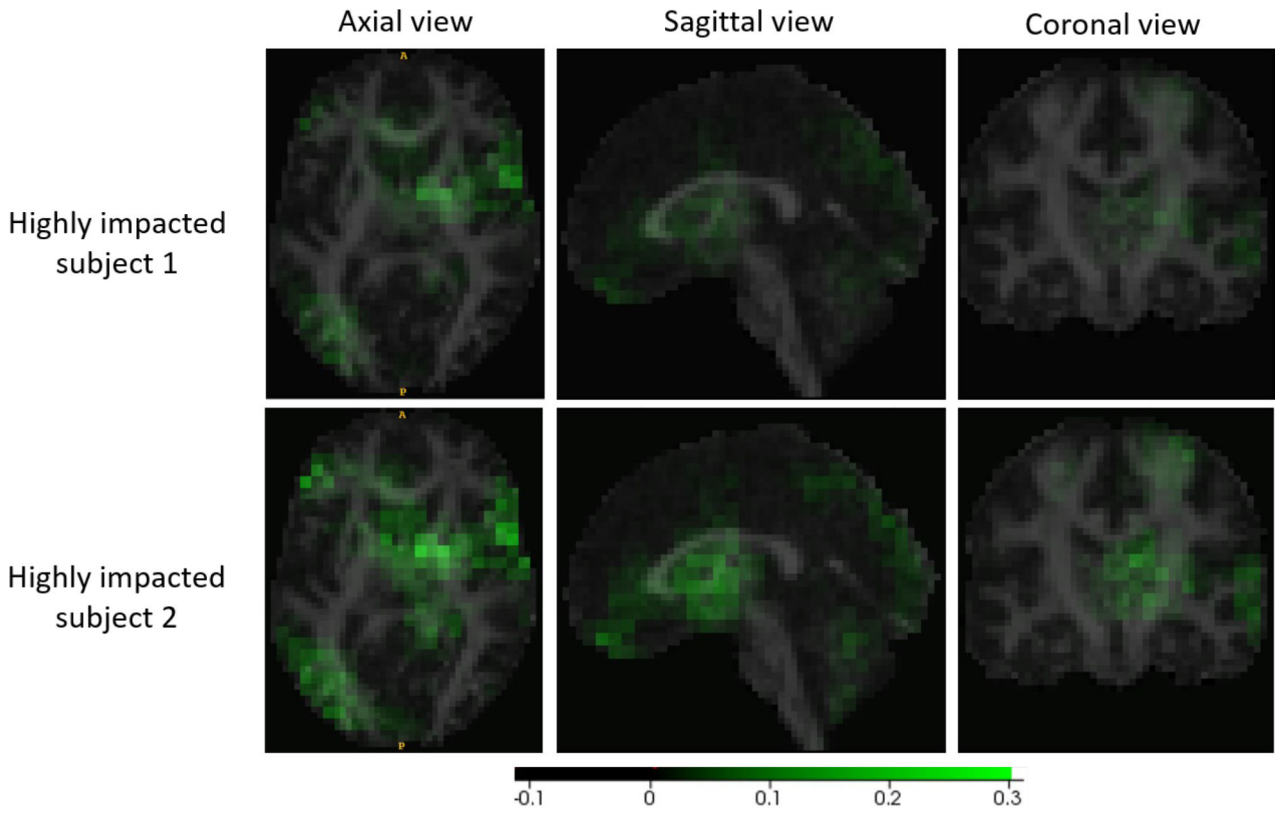


**Figure 3.**

Relative importance of regions used by the 3D CNN for the two most confident and correct high impact exposure player predictions. Row 1 is the most confident (94%) high impact exposure prediction and row 2 is the second most confident (86%). Overlay (green) shows the region importance in green, while the underlay, the FA map at post-season, provides anatomical contrast. The model uses a preponderance of regions near the center of mass of the head as well as regions on opposite sides of the brain.

**Table 1.**

Architecture of the proposed 3DCNN

| Layer Name | Layer Type | Filter Size | Filter Number | Feature Map Size | Padding | Activation |
|---|---|---|---|---|---|---|
| In | Input | | | 48×60×52 | | |
| B11 | Batch Normalization | - | - | 48×60×52 | - | - |
| C11 | 3D Convolution | 3×3×3 | 8 | 48×60×52×8 | Stride 1×1×1 | PReLU |
| C12 | 3D Convolution | 3×3×3 | 8 | 48×60×52×8 | Stride 1×1×1 | PReLU |
| M1 | 3D Maxpooling | 2×2×2 | - | 24×30×26×8 | - | - |
| C21 | 3D Convolution | 3×3×3 | 16 | 24×30×26×16 | Stride 1×1×1 | PReLU |
| C22 | 3D Convolution | 3×3×3 | 16 | 24×30×26×16 | Stride 1×1×1 | PReLU |
| M2 | 3D Maxpooling | 2×2×2 | - | 12×15×13×16 | - | - |
| C31 | 3D Convolution | 3×3×3 | 32 | 12×15×13×32 | Stride 1×1×1 | PReLU |
| C32 | 3D Convolution | 3×3×3 | 32 | 12×15×13×32 | Stride 1×1×1 | PReLU |
| M3 | 3D Maxpooling | 2×2×2 | | 6×6×7× 32 | | |
| F1 | Flatten | - | - | 8,064 | - | - |
| F2 | Dense | - | - | 32 | - | PReLU |
| D1 | Dropout(0.3) | - | - | 32 | - | - |
| F3 | Dense | - | - | 8 | - | PReLU |
| Out | Dense (Output) | - | - | 2 | - | Softmax |

**Table 2.**

Performance of the proposed 3DCNN on the train and held out test set. The model is trained on 48 subjects and tested on 12 subjects.

| | Sensitivity(%) | Specificity(%) | ROC AUC(%) | F1-score(%) |
|---|---|---|---|---|
| Train Scores | 100.00 | 100.00 | 100.00 | 100.00 |
| Test Scores | 100.00 | 71.43 | 85.71 | 83.33 |