Detecting Carpal Tunnel Syndrome with Machine Learning on Ultrasonography

Tianyi Li

Date: August 29, 2020

Contents

1 Introduction

Carpal tunnel syndrome (CTS) is a universal entrapment neuropathy that causes compression and thus symptoms at the carpal tunnel. In order to diagnose CTS, electrodiagnostic tests (EDTs) are performed on the potential patients, while the tests' being invasive and time-consuming is somehow intolerant by patients[1]. To make the case worse, it's still controversial whether EDTs are accurate and reliable enough for the diagnoses, as several studies have argued[3].

In the 21^{st} century, it is more likely for the radiologists to detect human's body, thanks to the development of image techniques. Among all the techniques, ultrasonography (US) outperforms the others in the detection of CTS. It's capability of returning high-resolution images of one's carpal tunnel along the transverse or longitudinal axis using untrasound can effectively assist physiologists to examine whether a patient suffers CTS or not by directly inspecting the morphological features of the carpal tunnel as well as its surroundings.

Although everything seems work pretty well, certain problems still exist. First of all, to objectively diagnose whether a subject suffers CTS, radiologists set cut points of certain characteristics of the carpal tunnel retracted from the image as thresholds. However, simple partitions don't seem to possess high enough sensitivity or specificity that can be widely accepted[4]. Secondly, the whole process may take radiologists plenty of time, which potentially influences the diagnostic efficiency. To deal with the present case, a few studies suggest applying machine learning, which not only performs extraordinarily in classification tasks but also saves time for the physiologists as the computer can reach the result automatically. However, the previous studies either didn't receive good enough results, or emphasize on features other than morphology, say the elasticity parameters in the training process. From a new perspective, we would like to utilize morphological parameters of the carpal tunnel, especially the median nerve, to establish the models on the basis of ultrasonography.

2 Experimental Methodology

2.1 Subjects and Data

...REDACT

2.1.1 Image Processing

For the image processing part, we will be aiming at segmenting the MN from the original sonography image in order to extract its morphological parameters. Such work can be done manually, however the accuracy of segmentation and efficiency of the process can not be guaranteed. As a result, algorithms for segmenting the MN are to be found. As we segment the image, prominent morphological parameters, such as cross sectional area(CSA), perimeter, AP ratio, FR, etc. will be extracted. Also, since we may need plenty of images to fulfill the need of the following training part, data augmentation is to be done to expand the sample space.

2.2 Traditional Machine Learning Models

For the research, we firstly took advantage of several traditional machine learning(TML) models that have already been proved effective and prominent in classification tasks. Specifically, we applied K-nearest Neighbor(KNN), Logistic Regression(LR), Random Forest(RF), Gradient Boosting(GB), and Support Vector Machine(SVM).

2.2.1 Data Embedding

Specifically, we applied all three levels of sonographs from one subject's carpal tunnel. For each level, we measured the median nerve's

- CSA: cross sectional area (mm^2)
- perimeter (mm)
- AP ratio: the ratio of area square to perimeter (mm^3)
- FR: flattening ratio, namely the ratio of minor to major diameter

assuming that the contour of one's median nerve follows a elliptical shape. As Figure 1 shows, one can in advance segment out the median nerve, and then look for the morphological measurements by applying the following formula:

$$CSA = a * b * \pi$$
$$Perimeter = 4 * ((a + b) - (4 - \pi) * a * b/(a + b))$$
$$APratio = \frac{CSA^2}{Perimeter}$$
$$FR = \frac{b}{a}$$

where a and b are respectively the radius of minor and major axis of the ellipse. Here the approximation formula for perimeter has a 0.003 uncertainty.

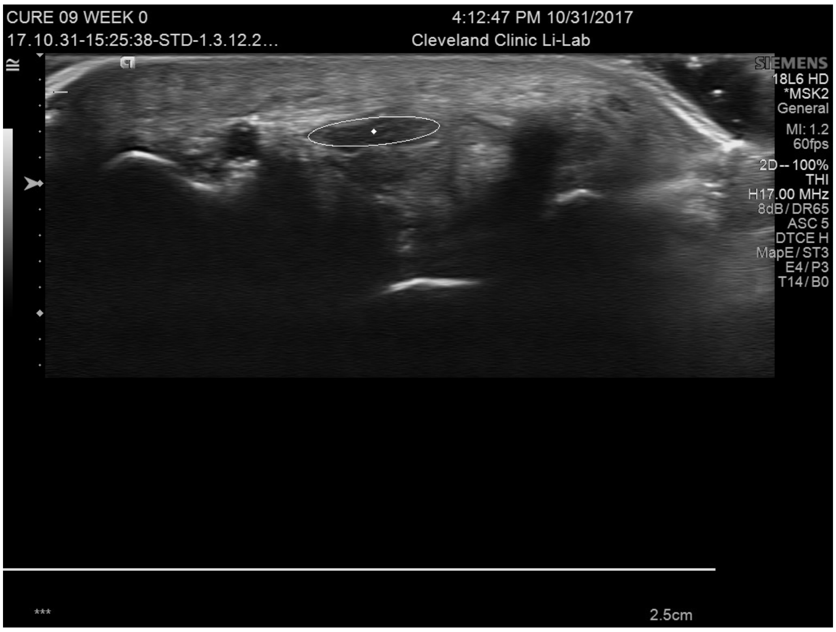

Figure 1: Contour segmentation for one's median nerve.

Practically, we apply $\pi = 3.14$. The distribution of each feature on three levels are shown below in Figure 2. After extracting morphological data, normalization is applied to the feature vectors.

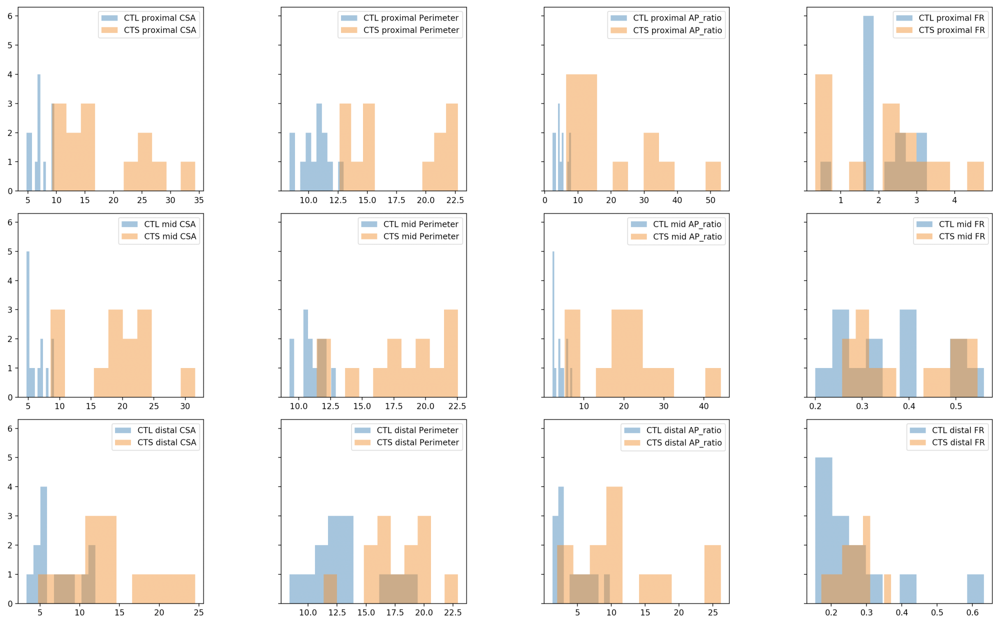

Figure 2: Feature Visualization

2.2.2 K-nearest Neighbor

K-nearest Neighbor, or KNN, is one of the most basic TML models in classification tasks. KNN adopted the perceptual intuition that for samples which share similar features or attributes, they are more likely to be classified into the same category. Based on this algorithm, for any test case under classification, KNN will find k samples with known category that are nearest to the test case in the feature vector space, and determine the category test sample falls in based on those k samples. In the research, we tuned and choose the hyper-parameter k = 3, namely the result of classification for each test case will base on three samples nearest to it.

2.2.3 Logistic Regression

Logistic Regression, or LR, is another basic TML model that works well in classification tasks which applies logistic function. For each test case, it will be classified as positive when the possibility is larger than 0.5, while as negative when possibility is less than 0.5. In the research, we tuned the hyper-parameter inverse regularization strength = 1.0, under which the model reaches the best performance.

2.2.4 Random Forest

Random Forest, or RF, is an ensemble classifier that gathers the predictions from a multiple number of decision trees (DTs). By doing so RF can somehow avoids overfitting which possibly results from one decision tree, meanwhile improves the accuracy of prediction. In the research, we tuned several hyper-parameters, respectively the criterion for each classification in decision trees, number of trees and maximum depth of each tree. It turns out that RF model will reach the best performance when applying the Gini impurity as criterion, 100 trees to ensemble and each tree with maximum depth of 10.

2.2.5 Gradient Boosting

Gradient Boosting, or GB, is another ensemble classifier which uses the predictions from multiple trees. However, unlike RF algorithm, trees in GB are trained in the negative gradient direction. By such method each tree can be trained more effectively and cost less. In the research, we tuned the hyper-parameter learning rate = 0.1, under which GB can reach the best performance.

2.2.6 Support Vector Machine

Support Vector Machine, or SVM, has been proved being one of the most prominent models in classfication tasks. In short, SVM sets up a boundary in the feature vector space during the training process. For binary classification tasks, test cases on one side of the boundary will be determined belonging to the same category. As a result, the training process of SVM is to maximize the distance from the boundary to the sample points adjacent to it. In the research, we tuned the hyper-parameter inverse regularizaion strength C = 1.0, under which circumstance the model reaches the best performance.

2.3 Neural Network Models

After applying TML models, we applied neural network models for further researches on the platform of Keras. Here we applied convolutional neural network to work out the classification.

2.3.1 Data Embedding

As mentioned above, data are presented in forms of sonographs, which are composed of grayscale pixels with values ranging from 0 to 255. However, since our focus is on the median nerve, instead of directly inputting the sonographs we instead base our models on the masks of the sonographs. A mask of a figure is a method to highlight important information and to exclude useless parts. Practically in our research, we colors the area of the median nerve while leave other parts uncolored, by which means we splits out the median nerve. On the basis of masks, it is more likely for the neural network to pick up information that is effective for classification.

2.3.2 Data Augmentation

Data augmentation is one of the effective ways to enlarge the sample space in neural network training. In short, data can be modified through translation, rotation, etc. so that the model will receive much more data inputs than the amount of original dataset.

2.3.3 Convolutional Neural Network

Convolutional Neural Network, or CNN, applies sliding windows and kernels to integrate n dimentional vectors into smaller ones via convolutional computation. CNN is one of the most prominent NN models used in the field of computer vision, thanks to its capability of reception and conform with figures.

Aside from the convolutional neural layer, the network is also composed of a multiple pooling layers, flattening layer, dropout layers and fully connected layers. After trying out different combinations and arrangements, we found that the best performance results from the following pipeline.

Layer	Mask	Conv2D	MaxPool2D	Conv2D	MaxPool2D	Flattening	Dropout
Size	(200, 200)	(198, 198)	(99,99)	(97, 97)	(48,48)	147456	147456

Table 1: CNN pipeline

3 Results

3.1 Traditional Machine Learning Performance

We select five traditional models, respectively K-nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), and Support Vector Machine (SVM). We applied a 13-fold validation process, and the results of average cross-validation scores are shown in Table 2.

Model	Avr. CV Score
KNN	1.00
LR	1.00
RF	1.00
GB	0.88
SVM	1.00

Table 2: Models' Average Cross-validation Score

On the basis of a relatively small sample size, the results are somehow acceptable, while still need great improvements for real life medical assistance. After further researching on the results, we found that the huge majority of mis-classifications took place on data with strange morphological features, say from CTS but with a rather small CSA.

Among all the models, SVM performs the best. Expect for the edge samples that all models fail to correctly categorize, SVM successfully identify nearly all other common subjects of CTS and CTL. We can interpret the performance of SVM from it's mathematical feature. Namely, SVM focuses more on the data points on or near the classification boundary(which are called support vectors), while less focuses on the points faraway. As one can imagine, points with strange features are commonly far from the classification boundary, and thus they don't affect the construction of SVM. Although SVM cannot identify those edge cases, it performs perfectly on other data. Comparatively speaking, other models will more or less be influenced by those edge data, leading to a lower accuracy. This is also verified by the fact that under a larger process of cross validation, one will observe that models other than SVM will make wrong estimates on common cases at a certain rate, on which SVM still holds a high-enough accuracy.

3.2 Future Research

As mentioned, the models are still lack of power in identifying edge cases. To deal with the problem, there are two main future improvements we may take. First, enlarging the dataset may do a great favor. Second, deep learning can be applied, since it can by itself figure out certain hidden features directly from the sonographs that people possibly never notice.

4 Bibliography

[1] Ardakani, A.A., A. Ahmadreza, S. Bhatt, N.J Bureau, A. Tahmasebi, U R. Acharya and A. Mohammadi. (2020). Diagnosis of Carpal Tunnel Syndrome: A Comparative Study of Shear Wave Elastography, Morphometry and Artificial Intelligence Techniques. Pattern Recognition Letters. 10.1016/j.patrec.2020.02.020.

[2] Buchberger, W., W. Judmaier, G. Birbamer, M. Lener, and C. Schmidauer. (1992). Carpal tunnel syndrome: Diagnosis with high-resolution sonography. AJR. American journal of roentgenology. 159. 793-8. 10.2214/ajr.159.4.1529845.

[3] El Miedany, Y.M., S.A. Aty and S. Ashour. (2004). Ultrasonography versus nerve conduction study in patients with carpal tunnel syndrome: Substantive or complementary tests?. Rheumatology (Oxford, England). 43. 887-95. 10.1093/rheumatology/keh190.

[4] Wong, Shiu Man, James F. Griffith, Andrew C.F. Hui, Sing Kai Lo, Michael Fu, and Ka Sing Wong. (2004). Carpal Tunnel Syndrome: Diagnostic Usefulness of Sonography. Radiology. 232. 93-9. 10.1148/radiol.2321030071.