Assessment of Stroke Risk Based on Morphological Ultrasound Image Analysis with Conformal Prediction

Antonis Lambrou[1], Harris Papadopoulos[2,1], Efthyvoulos Kyriacou[2,3], Constantinos S. Pattichis[3], Marios S. Pattichis[3,4], Alexander Gammerman[1], and Andrew Nicolaides[3,5]

[1] Computer Learning Research Centre, Royal Holloway, University of London, UK
{A.Lambrou,A.Gammerman}@cs.rhul.ac.uk
[2] Computer Science and Engineering Department, Frederick University, Cyprus
{H.Papadopoulos,E.Kyriacou}@frederick.ac.cy
[3] Computer Science Department, University of Cyprus, Nicosia, Cyprus
{anicolai,pattichi}@ucy.ac.cy
[4] Electrical and Computer Engineering Department, University of New Mexico
pattichis@ece.unm.edu
[5] Cyprus Institute of Neurology and Genetics

Abstract. Non-invasive ultrasound imaging of carotid plaques allows for the development of plaque image analysis in order to assess the risk of stroke. In our work, we provide reliable confidence measures for the assessment of stroke risk, using the Conformal Prediction framework. This framework provides a way for assigning valid confidence measures to predictions of classical machine learning algorithms. We conduct experiments on a dataset which contains morphological features derived from ultrasound images of atherosclerotic carotid plaques, and we evaluate the results of four different Conformal Predictors (CPs). The four CPs are based on Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Naive Bayes classification (NBC), and k-Nearest Neighbours (k-NN). The results given by all CPs demonstrate the reliability and usefulness of the obtained confidence measures on the problem of stroke risk assessment.

keywords: Conformal Prediction, stroke risk assessment, ultrasound images, confidence measures, carotid plaques.

1 Introduction

Visual classification of high-resolution ultrasound has made the non-invasive visualisation of the carotid bifurcation possible, and has thus been used in the study of arterial wall changes. Clinical applications of carotid bifurcation ultrasound include: i) identification and grading of stenosis of extracranial carotid artery disease often responsible for ischemic strokes, Transient Ischemic Attacks

(TIAs) or Amaurosis Fugax (AF); ii) follow-up after carotid endarterectomy; iii) evaluation of pulsatile neck mass; iv) investigation of asymptomatic neck bruits where severe internal carotid artery stenosis is used as a predictive factor for future stroke; v) cardiovascular risk assessment where the presence of carotid bifurcation atherosclerotic plaques is associated with increased cardiovascular mortality. During the last 20 years, the introduction of computer aided methods and image standardisation has improved the objective assessment of carotid plaque echogenicity and heterogeneity [1], and has largely replaced subjective assessment that had been criticized for its poor reproducibility [2].

In this work, we propose the use of Conformal Prediction for assigning reliable confidence measures to the classification of plaques into symptomatic or asymptomatic, based on ultrasound images. To our knowledge, no other method which provides any type of confidence measures has been used before on this problem. Unlike most classification techniques, Conformal Predictors (CPs) can provide predictive regions which guarantee, under the i.i.d. assumption, that the error rate of the predictive regions will be bounded by a desirable significance level. Several machine learning classifiers can be incorporated into the Conformal Prediction framework. Work in [3,4,5,6,7,8] has been conducted for building and evaluating CPs using Support Vector Machines, k-Nearest Neighbours, Artificial Neural Networks, and Ridge Regression. Moreover, the Conformal Prediction framework has been applied to medical diagnostic problems with success for breast cancer [9], acute abdominal pain [10], ovarian cancer [11], and leukemia diagnosis [12].

We experiment on a real-world dataset which consists of morphological features derived from ultrasound images of atherosclerotic carotid plaques [13]. We apply the Conformal Prediction framework using four different classifiers: Artificial Neural Network (ANN); Support Vector Machine (SVM); Naive Bayes Classification (NBC); and k-Nearest Neighbours (k-NN). We compare the results and we show the reliability and practicality of the confidence measures obtained for the classification of atherosclerotic carotid plaques.

The rest of the paper is structured as follows. In section 2, we describe the data used, we give an overview of the Conformal Prediction framework, and explain how we have transformed four machine learning classifiers into CPs. In section 3, we describe our experimental settings and we give the results we have achieved. In section 4, we conclude and we outline our plans for future work.

2   Material and Methods

2.1   Atherosclerotic Carotid Plaque Data

A total of 274 carotid plaque ultrasound images associated with retinal or hemispheric symptoms (33 stroke, 60 TIA, and 44 AF) were used in this work. Patients with cardioembolic symptoms or distant symptoms (>6 months) were excluded from the study. Asymptomatic plaques were truly asymptomatic if they had never been associated with symptoms in the past, or symptomatic if they had been associated with retinal or hemispheric symptoms (Stroke, TIA or AF).

The ultrasound images were collected in the Irvine Laboratory for Cardiovascular Investigation and Research, Saint Mary's Hospital, UK, using an Advanced Technology Laboratories (ATL model HDI 3000 - Seattle, USA) duplex scanner with a linear broadband width 4-7 MHz (multifrequency) transducer, at a resolution of 20 pixels/mm. The gray scale images (gray levels 0-255) were normalized manually by adjusting the image linearly. The plaque identification and segmentation tasks are quite difficult and were carried out manually by a physician or vascular ultrasonographer who are experienced in scanning, both actions are described in [14]. The morphological features derived from the images are motivated from the need to study the structure of the plaque. In this work, we have used the group of L-images as described by the Multilevel binary morphological analysis in [13]. This group gave the best accuracy results.

2.2 Conformal Prediction

The Conformal Prediction framework provides a way for assigning reliable confidence measures to predictions, based on an underlying machine learning algorithm. Typically, we are given a training set of the form $\{(x_1, y_1), ..., (x_n, y_n)\}$, where x_i is a vector of real-valued attributes and $y_i \in \{Y_1, Y_2, ..., Y_c\}$ is a label given to the instance x_i. For a new instance x_{n+1}, we intend to predict the label y_{n+1} (i.e. the class of the instance). In order to make a prediction, we assume all possible classes $Y_h \in \{Y_1, Y_2, ..., Y_c\}$ for the new instance, and we test for each one how likely the prediction is of being correct. In order to test each assumption, we append the new instance x_{n+1} in our training set together with the assumed class Y_h, and we train the underlying machine learning algorithm on the extended training set

$$\{(x_1, y_1), ..., (x_{n+1}, Y_h)\}. \tag{1}$$

We then calculate a non-conformity score for each instance in (1). A non-conformity score indicates how different (or strange) an instance x_i is for its label y_i, compared to the other instances in (1). In section 2.3, we explain how we have modified four underlying algorithms in order to generate non-conformity scores. Once we have non-conformity scores for the instances in (1), we measure how likely the extended training set is of being i.i.d., using the p-value function

$$p(Y_h) = \frac{\#\{i = 1, ..., n+1 : a_i \geq a_{n+1}\}}{n+1}, \tag{2}$$

which compares the non-conformity score a_{n+1} of (x_{n+1}, Y_h) with all the other non-conformity scores. We call the output of this function the p-value of the class Y_h. For the wrong prediction Y_h, we expect that a_{n+1} will be relatively higher than most of the non-conformity scores. In such cases, we will get low p-values, whereas for the correct prediction we expect a higher p-value.

For the true label of x_{n+1}, the p-value function in (2) satisfies the following property for all probability distributions P, and for any significance level ϵ:

$$P(p(y_{n+1}) \leq \epsilon) \leq \epsilon. \tag{3}$$

Table 1. Example of a certain and an uncertain prediction at 95% confidence

Instance	x_1	x_2
$p(Y_1)$	0.8623	0.1920
$p(Y_2)$	0.0145	0.3768
Actual label	Y_1	Y_2
Predictive region for $\epsilon = 0.05$	$\{Y_1\}$	$\{Y_2, Y_1\}$

The property describes that when the given training set contains i.i.d. instances, the probability of the p-value of the training set to be less than or equal ϵ, is less than or equal ϵ. Consequently, we may output a set of possible predictions (i.e. a predictive region), which contains all the predictions with p-values greater than the significance level ϵ. Moreover, we always include the highest prediction in order to ensure that the predictive region will contain at least one prediction:

$$S = \{Y_h : p(Y_h) > \epsilon\} \cup \left\{ \arg \max_{h=1,...,c} (p(Y_h)) \right\}. \qquad (4)$$

Because of the property in (3), the probability of each set S not containing the correct prediction will be less than or equal to ϵ. As a result, the error of the predictive regions will be bounded to ϵ, and thus we can say that we have $1 - \epsilon$ confidence in our predictions. Alternatively, the CP may output a single prediction, which is the prediction with the highest p-value, complemented with a confidence measure, which is one minus the second highest p-value, and a credibility value which is the p-value of the prediction. The confidence measure shows how likely the output classification is of being correct, compared to all other possible classes. The credibility value gives an indication of how suitable the training set is for classifying the current instance (i.e. if the credibility value is very low, then the training set is not i.i.d. or the current instance is strange).

In Table 1, we give an example of a predictive region which contains a single label (certain prediction) and a predictive region which contains both labels for 95% confidence level. For instance x_1, the second p-value is 0.0145, which is less than the significance level of 0.05. Therefore, we can discard the second label at 95% confidence, and give a certain prediction which is the label that gives the highest p-value. In contrast, for instance x_2, the second largest p-value is 0.1920 and is greater than the significance level of 0.05. In this case, we cannot discard the second largest p-value at 95% confidence, and thus we have an uncertain predictive set, which contains both possible labels. Nevertheless, if we decrease confidence at 80.80% (or lower), we then have a certain prediction.

2.3   Non-conformity Measures

We describe how we derive non-conformity measures from four classical machine learning algorithms. Specifically, we give non-conformity measures for Artificial Neural Networks, Support Vector Machines, the Naive Bayes Classifier, and k-Nearest Neighbours.

Artificial Neural Networks. Artificial Neural Network (ANN) classifiers are usually trained on a training set to re-adjust the weights of the connections between the units inside the network. The output layer of a neural network has a unit o_j for each possible class, and given an instance x_i we predict the class Y_j corresponding to the unit which gives the highest value. We expect that, the more conforming an instance is for its class, the higher the corresponding o_j value would be. As proposed in [3], we can build a CP based on ANNs (ANN-CP), using the non-conformity measure

$$\alpha_i = 1 - o_t, \tag{5}$$

for any (x_i, y_i) where $y_i = Y_t$. Alternatively, we can use the following non-conformity measure which is again defined in [3]:

$$\alpha_i = \frac{\max_{j=1,\dots,c:j\neq t} o_j}{o_t}. \tag{6}$$

That is, we use the maximum of the output units which do not correspond to the label of the given instance as the numerator, since a higher value from those units would also indicate a more strange instance, and would give a higher non-conformity score when divided by o_t.

Support Vector Machines. Support Vector Machines (SVMs) identify boundary instances for each class, and fix a separating hyperplane that maximises the margin between them. For the purpose of building a CP using SVM (SVM-CP), we use the distance of each instance from the separating hyperplane, and the class it belongs to, in order to produce non-conformity scores. For $Y = \{-1, 1\}$, we use the non-conformity measure

$$\alpha_i = -y_i h(x_i), \tag{7}$$

where $h(x_i)$ is the output of the SVM for the given instance x_i. The output of $h(x_i)$ is negative if the instance belongs to class -1, and positive if it belongs to class 1. If the prediction is correct, then the further the instance is from the hyperplane, the less the non-conformity score will be. In contrast, if the prediction is incorrect, the non-conformity score will increase as the distance from the hyperplane increases.

Naive Bayes Classifier. The Naive Bayes Classifier (NBC) is named after Bayes' theorem, and the "naive" assumption of attribute independence. The classifier multiplies the probabilities of the attributes given their class, and outputs the probability of label y_i given instance x_i. We can use the output probability to define a non-conformity measure and build a CP based on NBC (NBC-CP):

$$\alpha_i = 1 - P(y_i|x_i). \tag{8}$$

Nearest Neighbours. The k-Nearest Neighbours (k-NN) method computes the distance of a test instance from the other instances that are provided in the training set, and finds its k nearest instances. The prediction of the algorithm is the class which is the majority of the k instances. In the case of building a CP based on k-NN (k-NN-CP), we use the distances of the k nearest instances to define a non-conformity measure. The simplest approach is to calculate the total of distances of the k instances that belong to the class of instance x_i, since the nearer the instance is to its class, the less strange it is. Nonetheless, for a more accurate non-conformity measure we also take into consideration the distances of the k nearest instances that belong to other classes, since the nearer the instance x_i is to other classes the more strange it is. We build our k-NN-CP using the non-conformity measure defined in [4,8]:

$$\alpha_i = \frac{\sum_{j=1,\ldots,k} s_{ij}}{\sum_{j=1,\ldots,k} o_{ij}}, \tag{9}$$

where s_{ij} is the jth shortest distance of x_i from the instances of the same class, and o_{ij} is the jth shortest distance of x_i from the instances of other classes.

3   Experiments

We have applied Principal Component Analysis (PCA) on the dataset and selected its 6 features which accounted for 98% of its variance. For evaluation, we have applied the Leave-One-Out (LOO) method. Both these choices were made in order to be able to compare our results with [13]. The ANN-CP was structured with one hidden layer consisting of 3 units, and the output layer consisting of 2 units (one for each class). All units had a sigmoid activation function. We have used a learning rate of 0.3 and a momentum rate of 0.2. The ANN was trained for 500 epochs with 10% validation set, which was used to stop training when the performance on the validation set was deteriorating. For the SVM-CP, we used a Radial Basis Function (RBF) kernel mapping with a spread parameter of 0.1, and for the k-NN-CP we set the parameter $k = 10$.

3.1   Results

In Table 2, we compare the accuracy achieved by the four CPs with their corresponding machine learning algorithms. The difference of accuracy of each method with the corresponding CP is not significant, as expected. We would like to highlight that our aim is not to improve the accuracy level of the classification task, but rather to produce more informative predictions. On average all methods have an accuracy of 70.48%, while the SVM provides the best accuracy which is 73.72% (reported in [13]). In the left part of Table 3, we compare the certainty rates of the CPs, for given confidence levels 95%, 85%, and 75%. The certainty rates are calculated as the rate of predictive regions that contain only a single label. We do this in order to measure the quality of the p-values provided by each CP, and therefore the efficiency of the confidence measures. Although the SVM

Table 2. Accuracy comparison of 4 classical algorithms and the corresponding Conformal Predictors

Method	Classifier	CP
ANN	71.53%	71.90%
SVM	**73.72%**	73.36%
NBC	67.15%	66.79%
k-NN	69.34%	70.07%

Table 3. Certainty and error rates for 3 levels of confidence

Confidence level	95%	85%	75%	Confidence level	95%	85%	75%
ANN-CP Certainty	**33.9%**	**66.4%**	**87.9%**	**ANN-CP Error**	4.7%	13.9%	23.0%
SVM-CP Certainty	18.6%	54.3%	85.0%	**SVM-CP Error**	4.7%	14.9%	24.8%
NBC-CP Certainty	21.9%	58.7%	81.7%	**NBC-CP Error**	4.7%	14.9%	24.8%
k-NN-CP Certainty	28.1%	63.8%	86.8%	**k-NN-CP Error**	4.7%	14.9%	24.8%

has the best accuracy, it is the ANN-CP that provides the best certainty rates. The ANN-CP gives a 33.9% of certain predictive regions with 95% confidence. This is due to the difficulty of the classification task, which is reflected by the low 73.72% accuracy. Nevertheless, a 33.9% of patients will have a certain prediction and the error of such predictions will be at most 5%. Given the difficulty of the task, the 33.9% of certainty is arguably a useful result. Moreover, as we decrease the confidence level, the certainty rates increase dramatically. For example, at 85% confidence, the certainty rate given by the ANN-CP has increased to 66.4%.

In the right part of Table 3, we provide the error rates of all CPs for confidence levels 95%, 85%, and 75%. The error rates are calculated as the rate of the predictive regions that did not contain the true label. We confirm the validity of our confidence measures, as the error rates are below the significance level.

4   Conclusion

The classification of symptomatic and asymptomatic atherosclerotic plaques is a crucial task as it can be used to predict the risk of stroke. The accuracy levels are low since plaques with clear symptomatic features might have not given an event for reasons that need to be investigated. In this work, we have applied the Conformal Prediction framework on four machine learning algorithms in order to assign reliable confidence measures to the recognition of symptomatic or asymptomatic plaques; thus assess the risk of stroke. Our results demonstrate the validity of the produced confidence measures and their practicality. We believe that our contribution is suitable for the classification of plaques, as valid confidence measures may increase the quality of the decision-making process.

In the future, we aim to improve our confidence measures using meta-learning methods, such as boosting and ensemble algorithms. Moreover, we would like to experiment with different non-conformity measures for optimising further the accuracy and confidence measures of our methods.

References

1. Belgaro, G., Nicolaides, A., Laurora, G., Cesarone, M., Sanctis, M.D., Incandela, L., Barsotti, A.: Ultrasound morphology classification of the arterial wall and cardiovascular events in a 6-year follow-up study. Arteriosclerosis, Thrombosis, and Vascular Biology 16(7), 851–856 (1996)
2. Nicolaides, A., Shifrin, E., Bradbury, A., Dhanjil, S., Griffin, M., Belcaro, G., Williams, M.: Angiographic and duplex grading of internal carotid stenosis: can we overcome the confusion? Journal of Endovascular Therapy 3(2), 158–165 (1996)
3. Papadopoulos, H.: Inductive conformal prediction, theory and application to neural networks. In: Fritzsche, P. (ed.) Tools in Artificial Intelligence, I-Tech., Vienna, Austria, pp. 315–330 (2008), `http://intechweb.org/downloadpdf.php?id=5294`
4. Proedrou, K., Nouretdinov, I., Vovk, V., Gammerman, A.: Transductive confidence machines for pattern recognition. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) ECML 2002. LNCS (LNAI), vol. 2430, pp. 381–390. Springer, Heidelberg (2002)
5. Saunders, C., Gammerman, A., Vovk, V.: Transduction with confidence and credibility. In: Proceedings of the 16th International Joint Conference on Artificial Intelligence, Los Altos, CA, vol. 2, pp. 722–726. Morgan Kaufmann, San Francisco (1999)
6. Vovk, V., Gammerman, A., Saunders, C.: Machine-learning applications of algorithmic randomness. In: Proceedings of the 16th International Conference on Machine Learning (ICML 1999), pp. 444–453. Morgan Kaufmann, San Francisco (1999)
7. Papadopoulos, H., Proedrou, K., Vovk, V., Gammerman, A.: Inductive confidence machines for regression. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) ECML 2002. LNCS (LNAI), vol. 2430, pp. 345–356. Springer, Heidelberg (2002)
8. Papadopoulos, H., Vovk, V., Gammerman, A.: Qualified predictions for large data sets in the case of pattern recognition. In: Proceedings of the 2002 International Conference on Machine Learning and Applications (ICMLA 2002), pp. 159–163. CSREA Press (2002)
9. Lambrou, A., Papadopoulos, H., Gammerman, A.: Evolutionary conformal prediction for breast cancer diagnosis. In: 9th International Conference on Information Technology and Applications in Biomedicine (ITAB 2009). IEEE, Los Alamitos (2009)
10. Papadopoulos, H., Gammerman, A., Vovk, V.: Confidence predictions for the diagnosis of acute abdominal pain. In: Iliadis, L., Vlahavas, I., Bramer, M. (eds.) Artificial Intelligence Applications & Innovations III. IFIP International Federation for Information Processing, vol. 296, pp. 175–184. Springer, Heidelberg (2009)
11. Bellotti, T., Luo, Z., Gammerman, A., Delft, F.W.V., Saha, V.: Qualified predictions for microarray and proteomics pattern diagnostics with confidence machines. International Journal of Neural Systems 15(4), 247–258 (2005)
12. Bellotti, T., Luo, Z., Gammerman, A.: Reliable classification of childhood acute leukaemia from gene expression data using confidence machines. In: Proceedings of IEEE International Conference on Granular Computing (GRC 2006), pp. 148–153 (2006)
13. Kyriacou, E., Pattichis, M.S., Pattichis, C.S., Mavrommatis, A., Christodoulou, C.I., Kakkos, S., Nicolaides, A.: Classification of atherosclerotic carotid plaques using morphological analysis on ultrasound images. Applied Intelligence 30(1), 3–23 (2009)
14. Langsfeld, M., Gray-Weale, A.C., Lusby, R.J.: The role of plaque morphology and diameter reduction in the development of new symptoms in asymptomatic carotid arteries. J. Vasc. Surg. 9, 548–557 (1989)