Calibrated Probabilistic Predictions for Biomedical Applications

Antonis Lambrou*†, Harris Papadopoulos*†‡, Alexander Gammerman†
*Frederick Research Centre, Nicosia, Cyprus
†Computer Learning Research Centre
Royal Holloway, University of London, UK
{A.Lambrou, A.Gammerman}@cs.rhul.ac.uk
‡Department of Computer Science and
Engineering, Frederick University, Cyprus
{H.Papadopoulos}@frederick.ac.cy

***Abstract*—Venn Prediction (VP) is a machine learning framework that can be used to develop methods that provide well-calibrated probabilistic outputs. Unlike other probabilistic methods, the VP framework guarantees validity under the assumption that the data are independently and identically distributed (i.i.d.). Well-calibrated probabilistic outputs are of great importance, especially in biomedical applications. In this work, we develop a new Venn Predictor based on the Sequential Minimal Optimisation (SMO) algorithm and we examine its application to two real-world biomedical problems. We demonstrate in our results that our method can provide calibrated probabilistic outputs for predictions without any loss of accuracy. Moreover, we compare the outputs of our method with the probability outputs of SMO with logistic regression.**

***Index Terms*—Venn Prediction, Probability outputs, biomedicine**

I. Introduction

Recent developments in the biomedical research domain have given rise to many applications in bio-computing science, especially in machine learning, where high-dimensional datasets can be modeled. Nevertheless, most machine learning methods do not provide probabilistic outputs for their predictions which are very important for this kind of applications. There are some methods that output probabilities, but these can sometimes be misleading and unreliable in cases where the data assumptions are incorrect, or when the task is difficult.

In this work, we propose the use of Venn Prediction for producing well-calibrated probabilistic predictions for biomedical problems. Venn Predictors (VPs) are machine learning algorithms that can provide reliable probability estimates, based on the only assumption that the data are independently and identically distributed (i.i.d.).

The Venn Prediction framework is an extension to the Conformal Prediction framework which was introduced in [1]. Conformal Predictors (CPs) provide reliable measures of confidence for their predictions based on the i.i.d. assumption. Several CPs have been developed based on various algorithms such as Support Vector Machines [2], Ridge Regression [3], k-Nearest Neighbours for classification [4] and for regression [5], Random Forests [6], and Genetic Algorithms [7]. The computational efficiency of CPs has also been greatly improved using Inductive Conformal Prediction (ICP) [8], as demonstrated when combined with Ridge Regression [9], k-Nearest Neighbours [10], and more recently with Neural Networks [11], [12]. The CP framework has been successfully applied to medical problems, such as breast cancer diagnosis [13], classification of leukaemia subtypes [14], early detection of ovarian cancer [15], and acute abdominal pain diagnosis [16].

Unlike CPs which provide confidence measures, VPs output probabilistic intervals for each classification. The Venn Prediction framework was also introduced in [1] where the interested reader can find a thorough description. Since then, VPs have been developed based on k-Nearest Neighbours [17], Nearest Centroid [18] and Neural Networks [19], [20]. Furthermore, a VP based on a binary SVM has been developed in [21], where it has been compared with Platt's method in the batch setting. An Inductive Venn Predictor (IVP) has also been introduced in [22]. In this work, we develop a VP based on the Sequential Minimal Optimisation (SMO) algorithm, and we apply our method on two real-world biomedical datasets: The Ecoli and Dermatology datasets, both available at the University of California, Irvine (UCI) Machine Learning Repository [28]. We conduct experiments with the SMO algorithm, SMO with logistic regression that outputs probability estimates, and our VP. We compare the results of all the algorithms, and we demonstrate the reliability of the probability estimates that are given by our method.

The rest of the paper is structured as follows. In section II, we describe the Venn Prediction framework and give details of how we have developed a VP based on SMO. In section III, we provide our experimental results on the Ecoli and Dermatology datasets. Finally, in section IV, we conclude and outline our future work.

II. Venn Prediction

The Venn Prediction framework provides a way for estimating valid probabilities based on the i.i.d. assumption. Typically, we have a training set[1] of the form $\{z_1, \ldots, z_{n-1}\}$, where

[1]The training set is in fact a multiset, as it can contain some examples more than once.

each $z_i \in Z$ is a pair (x_i, y_i) consisting of the object x_i and its classification y_i. For a new object x_n, we intend to estimate its probability of belonging to each class $Y_j \in \{Y_1, \ldots, Y_c\}$. The Venn Prediction framework assigns each one of the possible classifications Y_j to x_n and divides all examples $\{(x_1, y_1), \ldots, (x_n, Y_j)\}$ into a number of categories based on a *taxonomy*. A taxonomy is a sequence A_n, $n = 1, \ldots, N$ of finite measurable partitions of the space $Z^{(n)} \times Z$, where $Z^{(n)}$ is the set of all multisets of elements of Z of length n. We will write $A_n(\{z_1, \ldots, z_n\}, z_i)$ for the category of the partition A_n that contains $(\{z_1, \ldots, z_n\}, z_i)$. In the next subsection, we define a taxonomy based on the output of the SMO algorithm.

After partitioning the examples into categories using a taxonomy, the empirical probability of each classification Y_k in the category τ_{new} that contains (x_n, Y_j) will be

$$p^{Y_j}(Y_k) = \frac{|\{(x^*, y^*) \in \tau_{new} : y^* = Y_k\}|}{|\tau_{new}|}. \quad (1)$$

This is a probability distribution for the class of x_n. So after assigning all possible classifications to x_n we get a set of probability distributions $P_n = \{p^{Y_j} : Y_j \in \{Y_1, \ldots, Y_c\}\}$ that compose the multi-probability prediction of the VP. As proved in [1], these are automatically well calibrated, regardless of the taxonomy used.

The maximum and minimum probabilities obtained for each label Y_k amongst all distributions $\{p^{Y_j} : Y_j \in \{Y_1, \ldots, Y_c\}\}$, define the interval for the probability of the new example belonging to Y_k. We denote these probabilities as $U(Y_k)$ and $L(Y_k)$, respectively:

$$U(Y_k) = \max_{j=1}^{c} p^{Y_j}(Y_k) \quad (2)$$

$$L(Y_k) = \min_{j=1}^{c} p^{Y_j}(Y_k) \quad (3)$$

The VP outputs the prediction $\hat{y}_n = Y_{k_{best}}$, where

$$k_{best} = \arg\max_{k=1,\ldots,c} \overline{p(k)}, \quad (4)$$

and $\overline{p(k)}$ is the mean of the probabilities obtained for label Y_k amongst all probability distributions. The probability interval for $Y_{k_{best}}$ is $[L(Y_{k_{best}}), U(Y_{k_{best}})]$.

A. Venn Predictor with SMO

In this work, we use the Sequential Minimal Optimisation (SMO) [23] method as the underlying algorithm to divide examples into categories. We have used the implementation of the SMO which was developed in WEKA [24]. SMO efficiently solves the optimisation problem which arises during the training phase of Support Vector Machines (SVMs). In this version of SMO, multi-class problems are solved using pairwise classification (1-vs-1), and if logistic regression is used, the logistic models are built using pairwise coupling [25]. The VP taxonomy is simply based on the classification output of the SMO algorithm; the taxonomy categorizes the examples according to the classification given by the SMO. In order to build the label distributions as described in section II, we have used the transductive framework, in which for every test example we assume every possible label of the example and we add it in the training set. We train the SMO algorithm several times (one for each label assumption), and we build the label distributions based on the SMO taxonomy.

III. Experiments and Results

The SMO algorithm was used in our experiments in four variations: SMO classifier, SMO with Logistic Regression (SMOL), SMO with Feature Selection (SMO-FS), and SMO with Logistic Regression and Feature Selection (SMOL-FS). The FS method that we have used is the Correlation Based Feature Selection method (CBFS). The Venn Predictor algorithm was also used in four variations: SMO Venn Predictor (VENN-SMO), SMO Venn Predictor with Logistic Regression (VENN-SMOL), SMO Venn Predictor with CBFS (VENN-SMO-FS), and SMO Venn Predictor with Logistic Regression and CBFS (VENN-SMOL-FS).

We have conducted two sets of experiments: one in the offline setting and one in the online. In the offline setting, we have experimented with 10-fold cross validation, and we show the mean accuracy. For the algorithms which provide probabilistic outputs we also calculate the Brier Score (BS) which is defined as:

$$BS = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{c} (f(x_{ij}) - o_{ij})^2, \quad (5)$$

where $f(x_{ij})$ is the probabilistic output of the algorithm for example x_i and class j, and o_{ij} is set to 1 if example x_i belongs to class j, and 0 otherwise. For VPs, $f(x_{ij})$ is the mean of the probabilities obtained for class j. Additionally, we output the mean probability estimate, and for the VPs the mean probability interval.

In the online experiments we have selected SMOL-FS and our VENN-SMO-FS algorithms which provided the best results in the offline setting. The online experiments allow us to demonstrate the advantages of our VP over the other methods. In the online setting there is no initial training set. A test instance x_t is predicted by the algorithm and a probability estimate is given for the prediction. The test instance with its true label is then added to the training set which grows every time. In the results of the SMOL-FS algorithm, we graph the Cumulative Mean Probability (CMP) and the Cumulative Mean Accuracy (CMA) curves:

$$CMP(t) = \frac{1}{t} \sum_{i=1}^{t} \max_{j=1}^{c} f(x_i), \quad (6)$$

$$CMA(t) = \frac{1}{t} \sum_{i=1}^{t} Acc_i, \quad (7)$$

where t is the number of test examples that have been added to the training set, and $Acc_i = 1$ when the prediction for example x_i is correct and 0 otherwise.

For the VENN-SMO-FS algorithm we graph the Cumulative Mean Lower Probability (CMLP), the Cumulative Mean Upper

TABLE I
ATTRIBUTE INFORMATION OF THE ECOLI DATASET.

#	Short name	Description
1	Sequence Name	Accession number for the SWISS-PROT database
2	mcg	McGeoch's method for signal sequence recognition.
3	gvh	von Heijne's method for signal sequence recognition.
4	lip	von Heijne's Signal Peptidase II consensus sequence score.
5	chg	Presence of charge on N-terminus of predicted lipoproteins.
6	aac	score of discriminant analysis of the amino acid content of outer membrane and periplasmic proteins.
7	alm1	score of the ALOM membrane spanning region prediction program.
8	alm2	score of ALOM program after excluding putative cleavable signal regions from the sequence.

TABLE II
CLASS DISTRIBUTION OF THE ECOLI DATASET.

Short name	Description	#
cp	cytoplasm	143
im	inner membrane without signal sequence	77
pp	perisplasm	52
imU	inner membrane, uncleavable signal sequence	35
om	outer membrane	20
omL	outer membrane lipoprotein	5
imL	inner membrane lipoprotein	2
imS	inner membrane, cleavable signal sequence	2

Probability (CMUP), the Cumulative Mean Central Probability (CMCP), and the CMA curves:

$$CMLP(t) = \frac{1}{t}\sum_{i=1}^{t} L_i(Y_{k_{best}}), \tag{8}$$

$$CMUP(t) = \frac{1}{t}\sum_{i=1}^{t} U_i(Y_{k_{best}}), \tag{9}$$

$$CMCP(t) = \frac{1}{t}\sum_{i=1}^{t} \frac{U_i(Y_{k_{best}}) + L_i(Y_{k_{best}})}{2}. \tag{10}$$

A. Ecoli dataset

The Ecoli dataset contains 336 Ecoli proteins that are classified into 8 cellular localization sites [26]. In Table I we list the attributes of each protein, and in Table II we show the class distribution of the dataset. The mean accuracy achieved in [26] after a randomized 10-fold cross validation experiment on the same dataset was 81%.

1) Offline experiments: In Table III we list the accuracies of the SMO and SMO-FS algorithms with 10 different RBF spread parameters. The best mean accuracy achieved by both algorithms is 78.87% with RBF spread parameter set to 9.5.

In Table IV we show the results of the SMOL algorithm which outputs probability estimates. We show the mean accuracy, the Brier Score (BS), and the mean probability estimate of the algorithm using the best RBF parameter that is shown in Table III. Additionally, we include the results of the SMOL-FS algorithm.

TABLE III
10-FOLD CROSS VALIDATION ACCURACY RESULTS OF THE SMO AND SMO-FS ALGORITHMS ON THE ECOLI DATASET.

RBF	SMO Acc.	SMO-FS Acc.
5	77.98%	77.98%
5.5	77.98%	77.98%
6	78.27%	77.98%
6.5	77.98%	77.98%
7	77.68%	77.38%
7.5	78.27%	78.27%
8	78.27%	78.27%
8.5	78.27%	78.27%
9	78.27%	78.57%
9.5	**78.87%**	**78.87%**
10	78.57%	78.57%

In Table V we show the results of the SMO VPs on the Ecoli dataset. We show the mean accuracy, Brier Score (BS), the mean lower probability bound, and the mean upper probability bound. Remarkably, all VPs have achieved an accuracy of around 85%, which is significantly higher than the mean accuracy of the SMO algorithm. Moreover, the BSs of the VPs are significantly lower than those of the SMOL algorithm, especially in the case of the VENN-SMO-FS algorithm which gives a BS of 22.29%, a significant 10% difference than the worse 35.04% of the SMOL-FS algorithm. This demonstrates that the probability estimates of the VPs are much nearer to the real probabilities. From the results, we can also see how the mean accuracies always fall within the lower and upper bounds of the VPs.

TABLE IV
10-FOLD CROSS VALIDATION RESULTS OF THE SMOL AND SMOL-FS ALGORITHMS ON THE ECOLI DATASET.

Method	RBF	Acc.	BS	Mean Prob.
SMOL	9.5	77.38%	34.38%	72.01%
SMOL-FS	9.5	77.68%	35.04%	72.08%

2) Online experiments: In Figure 1 we graph the online results of the SMOL-FS algorithm on the Ecoli dataset. Here we demonstrate the problem of the SMOL-FS algorithm. Due to the difficulty of the task, the algorithm was not able to estimate well calibrated probabilities. As shown in the graph, the CMA lies at around 80% during the online experiment, while the probability estimates lie at around 90%. This difference of about 10% in the results can mislead to wrong judgment. In Figure 2 we show the online results of

TABLE V
10-FOLD CROSS VALIDATION RESULTS OF THE VENN-SMO, VENN-SMO-FS, VENN-SMOL, AND VENN-SMOL-FS ALGORITHMS ON THE ECOLI DATASET.

Method	RBF	Acc.	BS	Prob. Interval
VENN-SMO	**9.5**	**86.90%**	**22.14%**	**81.08% – 90.59%**
VENN-SMO-FS	9.5	86.90%	22.29%	81.04% – 90.49%
VENN-SMOL	9.5	85.42%	33.88%	22.19% – 91.05%
VENN-SMOL-FS	9.5	85.12%	33.88%	22.19% – 91.05%

our VENN-SMO-FS algorithm. The CMA is shown to fall within the bounds, and it also tends towards the CMCP curve. This result is expected and shows that one can rely on these probability estimates.

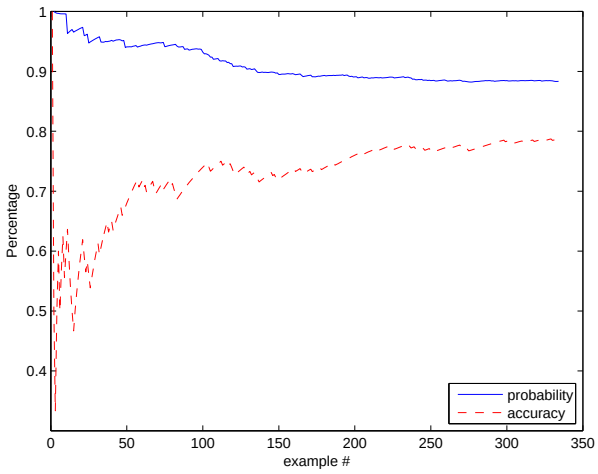

Fig. 1. Online experiment on the Ecoli dataset with the SMOL-FS algorithm. CMA and CMP curves are shown.

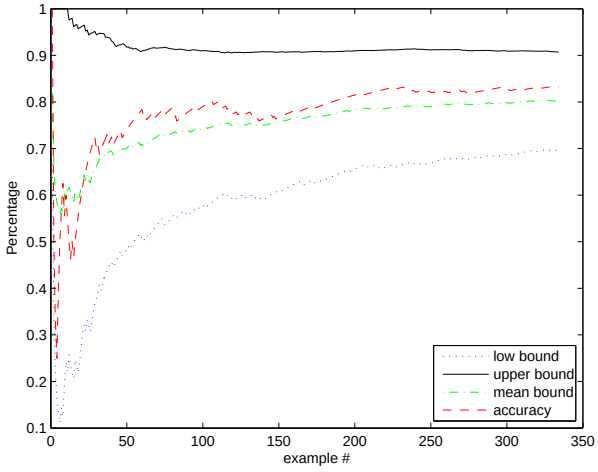

Fig. 2. Online experiment on the Ecoli dataset with our VENN-SMO-FS algorithm. CMA, CMLP, CMUP, and CMCP curves are shown.

B. Dermatology dataset

The Dermatology dataset contains 366 instances of patients with erythemato-squamous diseases [27]. The instances are classified into 6 diseases. In this work, 8 instances with missing values have been removed from the dataset. The reported accuracy in [27] is 96.2% after a 10-fold cross validation experiment. There are 34 attributes which are not listed here. The details about the attributes can be found in [28]. In Table VI we show the class distribution of the dataset.

TABLE VI
CLASS DISTRIBUTION OF THE DERMATOLOGY DATASET.

Class	Description	#
1	psoriasis	112
2	seboreic dermatitis	61
3	lichen planus	72
4	pityriasis rosea	49
5	cronic dermatitis	52
6	pityriasis rubra pilaris	20

TABLE VII
10-FOLD CROSS VALIDATION ACCURACY RESULTS OF THE SMO AND SMO-FS ALGORITHMS ON THE DERMATOLOGY DATASET.

RBF	SMO Acc.	SMO-FS Acc.
0.01	92.18%	95.81%
0.02	**93.30%**	**96.93%**
0.03	93.02%	96.37%
0.04	93.85%	96.93%
0.05	93.02%	96.93%
0.06	91.62%	96.37%
0.07	89.94%	96.37%
0.08	89.11%	96.65%
0.09	86.03%	96.65%
0.10	84.08%	96.37%

1) Offline experiments: In Table VII we list the mean accuracy of the SMO and SMO-FS algorithms on the dataset. The best mean accuracy achieved by SMO is 93.30%, and by the SMO-FS 96.93%. In Table VIII we include the results of the SMOL and SMOL-FS algorithms. The SMOL-FS algorithm has the best mean accuracy of 96.65% and BS 5.75%. In Table IX we show the results of the VPs on the Dermatology dataset. The best mean accuracy is achieved by the VENN-SMO-FS algorithm which is 96.93% and has a BS of 5.23% (slightly lower than the BS of SMOL-FS).

TABLE VIII
10-FOLD CROSS VALIDATION RESULTS OF THE SMOL AND SMOL-FS ALGORITHMS ON THE DERMATOLOGY DATASET.

Method	RBF	Acc.	BS	Mean Prob.
SMOL	0.02	90.50%	14.78%	89.26%
SMOL-FS	0.02	96.65%	5.75%	95.15%

2) Online experiments: In Figure 3 we show the online results of the SMOL-FS algorithm on the Dermatology dataset. In this case the SMOL-FS algorithm has perfomed quite well due to the fact that the dataset is easier for learning with the SMO algorithm. The CMA and CMP fall around 90%. Nevertheless, the probability bound given by the SMOL-FS algorithm can not always be trusted, as it was shown in Figure 1. In Figure 4 we show the online results of the VENN-

TABLE IX
10-FOLD CROSS VALIDATION RESULTS OF THE VENN-SMO, VENN-SMO-FS, VENN-SMOL, AND VENN-SMOL-FS ALGORITHMS ON THE DERMATOLOGY DATASET.

Method	RBF	Acc.	BS	Prob. Interval
VENN-SMO	0.02	93.30%	11.39%	78.75% – 97.82%
VENN-SMO-FS	**0.02**	**96.93%**	**5.23%**	**93.35% – 98.37%**
VENN-SMOL	0.02	90.78%	18.33%	50.44% – 96.71%
VENN-SMOL-FS	0.02	96.09%	7.98%	79.46% – 98.37%

SMO-FS algorithm on the Dermatology dataset. As it was stated previously, the CMA, CMLP, CMUP, and CMCP curves are shown. The accuracy is again well within the bounds as expected, and tends near the CMCP curve. Thus, the reliability of the probability estimates given by the Venn Predictor is demonstrated in the graph.

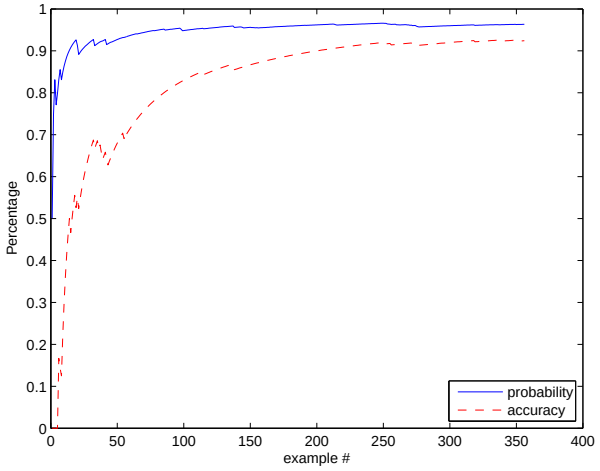

Fig. 3. Online experiment on the Dermatology dataset with the SMOL-FS algorithm. CMA and CMP curves are shown.

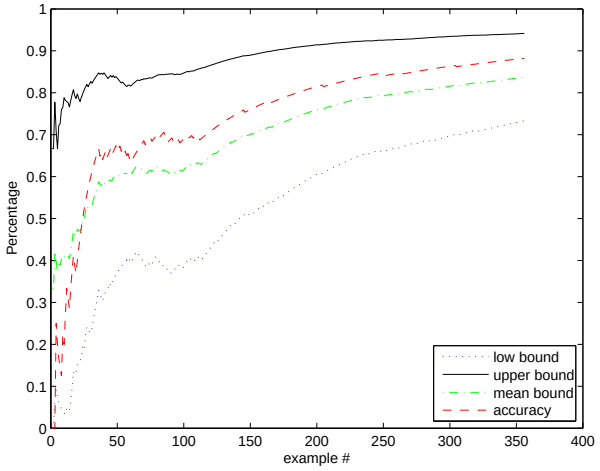

Fig. 4. Online experiment on the Dermatology dataset with out VENN-SMO-FS algorithm. CMA, CMLP, CMUP, and CMCP curves are shown.

IV. Conclusion

In this work, we have developed a VP for multi-class datasets based on the SMO algorithm. We have demonstrated the reliability of the probability estimates of Venn Predictors, and we have applied our method on two real-world biomedical problems. Unlike other probabilistic methods, Venn Predictors guarantee, under the assumption that the data are independently and identically distributed (i.i.d.), that the probability bounds will be well-calibrated.

In the future, our aim is to apply the proposed approach on biomedical problems where probabilistic predictions are of great importance. Furthermore, we aim to develop taxonomies that will increase the accuracy of our method.

Acknowledgment

This work was co-funded by the European Regional Development Fund and the Cyprus Government through the Cyprus Research Promotion Foundation "DESMI 2009-2010", research contract TPE/ORIZO/0609(BIE)/24 ("Development of New Venn Prediction Methods for Osteoporosis Risk Assessment").

References

[1] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*. New York, Springer, 2005.

[2] C. Saunders, A. Gammerman, and V. Vovk, "Transduction with confidence and credibility," in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, vol. 2. Los Altos, CA: Morgan Kaufmann, 1999, pp. 722–726.

[3] I. Nouretdinov, T. Melluish, and V. Vovk, "Ridge regression confidence machine," in *In Proceedings of the Eighteenth International Conference on Machine Learning*. Morgan Kaufmann, 2001, pp. 385–392.

[4] K. Proedrou, I. Nouretdinov, V. Vovk, and A. Gammerman, "Transductive confidence machines for pattern recognition," in *Proceedings of the 13th European Conference on Machine Learning (ECML'02)*, ser. Lecture Notes in Computer Science, vol. 2430. Springer, 2002, pp. 381–390.

[5] H. Papadopoulos, V. Vovk, and A. Gammerman, "Regression conformal prediction with nearest neighbours," *Journal of Artificial Intelligence Research*, vol. 40, pp. 815–840, 2011, URL http://dx.doi.org/10.1613/jair.3198. [Online]. Available: http://dx.doi.org/10.1613/jair.3198

[6] D. Dmitry and N. Ilia, "Prediction with confidence based on a random forest classifier," in *Artificial Intelligence Applications and Innovations - 6th IFIP WG 12.5 International Conference (AIAI 2010)*, vol. 339. Springer Boston, 2010, pp. 37–44. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-16239-8_8

[7] A. Lambrou, H. Papadopoulos, and A. Gammerman, "Reliable confidence measures for medical diagnosis with evolutionary algorithms," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 1, pp. 93 –99, 2011.

[8] H. Papadopoulos, "Inductive conformal prediction: Theory and application to neural networks," in *Tools in Artificial Intelligence*, P. Fritzsche, Ed. Vienna, Austria: I-Tech, 2008, ch. 18, pp. 315–330, URL http://www.intechopen.com/download/pdf/pdfs_id/5294.

[9] H. Papadopoulos, K. Proedrou, V. Vovk, and A. Gammerman, "Inductive confidence machines for regression," in *Proceedings of the 13th European Conference on Machine Learning (ECML'02)*, ser. Lecture Notes in Computer Science, vol. 2430. Springer, 2002, pp. 345–356.

[10] H. Papadopoulos, V. Vovk, and A. Gammerman, "Qualified predictions for large data sets in the case of pattern recognition," in *Proceedings of the 2002 International Conference on Machine Learning and Applications*, 2002, pp. 159–163.

[11] ——, "Conformal prediction with neural networks," in *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'07)*, vol. 2. IEEE Computer Society, 2007, pp. 388–395.

[12] H. Papadopoulos and H. Haralambous, "Reliable prediction intervals with regression neural networks," *Neural Networks*, vol. 24, no. 8, pp. 842–851, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2011.05.008

[13] A. Gammerman, I. Nouretdinov, B. Burford, A. Chervonenkis, V. Vovk, and Z. Luo, "Clinical mass spectrometry proteomic diagnosis by conformal predictors," *Statistical applications in genetics and molecular biology*, vol. 7, no. 2, 2008. [Online]. Available: http://dx.doi.org/10.2202/1544-6115.1385

[14] T. Bellotti, Z. Luo, and A. Gammerman, "Reliable classification of childhood acute leukaemia from gene expression data using confidence machines," in *Proceedings of IEEE International Conference on Granular Computing (GRC '06)*, 2006, pp. 148–153.

[15] A. Gammerman, V. Vovk, B. Burford, I. Nouretdinov, Z. Luo, A. Chervonenkis, M. Waterfield, R. Cramer, P. Tempst, J. Villanueva, M. Kabir, S. Camuzeaux, J. Timms, U. Menon, and I. Jacobs, "Serum proteomic abnormality predating screen detection of ovarian cancer," *Comput. J.*, vol. 52, no. 3, pp. 326–333, May 2009. [Online]. Available: http://dx.doi.org/10.1093/comjnl/bxn021

[16] H. Papadopoulos, A. Gammerman, and V. Vovk, "Reliable diagnosis of acute abdominal pain with conformal prediction," *Engineering Intelligent Systems*, vol. 17, no. 2-3, pp. 115–126, 2009.

[17] M. Dashevskiy and Z. Luo, "Predictions with confidence in applications," in *Machine Learning and Data Mining in Pattern Recognition*, ser. LNCS, P. Perner, Ed. Springer, 2009, vol. 5632, pp. 775–786.

[18] ——, "Reliable probabilistic classification and its application to internet traffic," in *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, ser. LNCS. Springer, 2008, vol. 5226, pp. 380–388.

[19] H. Papadopoulos, "Reliable probabilistic prediction for medical decision support," in *Proceedings of the 7th IFIP International Conference on Artificial Intelligence Appications and Innovations (AIAI 2011)*, ser. IFIP AICT, vol. 364. Springer, 2011, pp. 265–274.

[20] ——, "Reliable probabilistic classification with neural networks," *Neurocomputing*, 2012, to appear.

[21] C. Zhou, I. Nouretdinov, Z. Luo, D. Adamskiy, L. Randell, N. Coldham, and A. Gammerman, "A comparison of Venn Machine with Platt's method in probabilistic outputs," in *Proceedings of the 7th IFIP International Conference on Artificial Intelligence Appications and Innovations (AIAI 2011)*, ser. IFIP AICT, vol. 364. Springer, 2011, pp. 483–490.

[22] A. Lambrou, H. Papadopoulos, and A. Gammerman, "Reliable probability estimates based on support vector machines for large multiclass datasets," in *Artificial Intelligence Applications and Innovations (AIAI2012)*, L. Iliadis, I. Maglogiannis, H. Papadopoulos, K. Karatzas, and S. Sioutas, Eds. Springer, 2012, pp. 182–191.

[23] J. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods - Support Vector Learning*, B. Schoelkopf, C. Burges, and A. Smola, Eds. MIT Press, 1998.

[24] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, pp. 10–18, 2009.

[25] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," in *Advances in Neural Information Processing Systems*, M. I. Jordan, M. J. Kearns, and S. A. Solla, Eds., vol. 10. MIT Press, 1998.

[26] P. Horton and K. Nakai, "A probabilistic classification system for predicting the cellular localization sites of proteins," in *In Proceeding of the Fourth International Conference on Intelligent Systems for Molecular Biology*, 1996, pp. 109–115.

[27] H. A. Gvenir, G. Demirz, H. A. G. (corresponding, G. Demiroz, and N. Ilter, "Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals," *Artificial Intelligence in Medicine*, vol. 13, pp. 147–165, 1998.

[28] A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: http://archive.ics.uci.edu/ml