Dynamic combination of fusion methods by region of competence

line 1: 1st Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 2nd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 3rd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

Abstract*—This electronic document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. **CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract. **(***Abstract***)**

Keywords—component, formatting, style, styling, insert **(***key words***)**

I. Introduction

The use of classifiers committee is a sub-area that has been advancing over the years. Several forms have emerged, from the simplest as common combinations by average or majority vote, to more advanced combinations such as Neural Network or region of competence.

In this sense, several researches appear in order to discover the best ways to merge classifiers or to improve the results for specific cases. In several problems there are cases that algorithms, methods or ways of combining improve. Studies demonstrate various forms of these applications, such as the study by Dantas (2021).

When using a region of competence, it is possible to improve the combination of Classifiers, as only the most specialized in the region are selected. This is a way to improve results. But there are still problems, such as finding the best combination method for a new example, such as majority vote, average, maximum, among others.

A very explored dynamic factor is the Dynamic Selection of Committees. Several works show that the use of a set of classifiers for a given instance increases the predictive capacity of the system (WOODS; KEGELMEYER; BOWYER, 1997), (KO; SABOURIN; JR, 2008).

Thus, it is necessary to seek a better performance to maximize results. In this research, a study is carried out with a committee of classifiers in order to perform dynamic fusion, in order to choose the best combination method for a new instance.

A dynamic fusion of 10 Methods was carried out, plus the comparison of two methods of Voting and Weighted Weight, totaling the comparison of 12 methods with the one proposed in this research.

In order to evaluate the proposed method of this work, some approaches were performed for the dynamic selection of committees using the DESLIB library and the KNORA-E and META-DES combination methods.

II. COMMITTEES OF CLASSIFIERS

Research shows that the creation of committees of classifiers improves the results compared to a single classifier (KUNCHEVA, 2004), as it better combines the advantages and overcomes the individual limitations of the classifiers. Thus, they present diversity in their construction, which leads to a better ability to generalize compared to separate ones (DIETTERICH, 2000).

There are two categories of classifier ensemble architecture, modular and ensemble. This architecture is responsible for how they will interact within the system. In modular architecture, the problem is decomposed into subproblems, making each method an expert in one aspect of the problem. Finally, the obtained solutions are combined to determine the final output of the system (DANTAS, 2021).

The ensemble, the most commonly used and known, is composed of a set of parallel classifiers that receive patterns as input and send their outputs to a combination method, which is responsible for aggregating the final decision (KUNCHEVA, 2002) (BROWN et al. al., 2005). Thus, this method is more robust, as it uses different models to generate an output (BRAGA, 2005). Figure 1 shows this architecture.

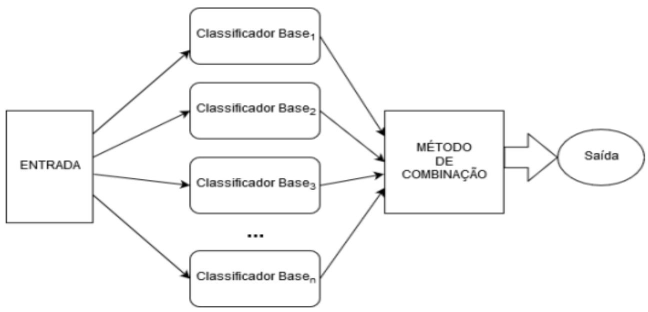

Fig 1. Structure Ensemble (DANTAS, 2021)

After data entry, each classifier will obtain an individual result for the test sets to which they are submitted. Finally, the combination method chosen combines the results into a single response. The construction of classifiers can vary through the choice of combination techniques, the distribution of different databases or even through training with different attributes. This is responsible for generating diversity in the classification system.

Another necessary definition is how the combination of the results of each classifier will be performed. In the literature, the combination strategies discussed are: selection and fusion (CANUTO et al., 2007).

In the fusion it is assumed that the values of each base classifier have knowledge about the entire feature space. The value selection techniques assume that each classifier is an expert in particular characteristics of the database. Based on the instance characteristics this technique chooses the classifier.

To perform the combination of the results obtained by each classifier that composes the committee, it is possible to use one of several value fusion techniques. Some of them are: Majority Vote, Average, Maximum, Minimum, Geometric Average, among others. Other more robust techniques that can be used are: Artificial Neural Network, Naive Bayes, Edge, among others.

A. Dynamic selection of committees

To perform the dynamic selection of committees there is the method DES (Dynamic Selection of Committees), which selects a subset of classifiers to classify the test instance. For this, techniques are applied to delimit the initial set of classifiers, to select a subset of classifiers that can better contribute to the classification of a test instance.

In this Research, the KNORA committee dynamic selection methods, through the KNORA-Eliminate variation, and the META-DES will be used.

KNORA

KNORA (KO; SABOURIN; JR, 2008) seeks to find the best subset of classifiers for a given test instance. When classifying a test sample, the K-nearest neighbors are selected from the validation set.

The subset of classifiers that will be used to classify the test instance is defined by discovering which classifiers from an initial set correctly classify the selected K-neighbors. Each test sample repeats the process so that each sample will be ranked using a committee specific to it.

In KNORA-Eliminate (KO; SABOURIN; JR, 2008) the selection, in a validation set, of the nearest K-neighbors of the instance to be classified begins. The basic idea is to use a set of classifiers formed only by the classifiers that correctly classify all K-nearest neighbors. If no classifier is able to correctly classify all K-neighbors, the value of K is decremented until at least one classifier can be added to the committee that will be responsible for classifying the test sample. This operation is shown in Figure 2.

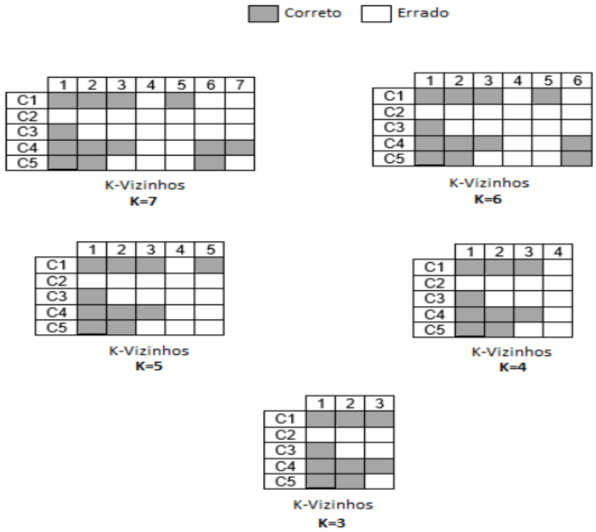

Fig 2. Example of how the KNORA-E works (DANTAS, 2021)

META-DES

META-DES (CRUZ et al., 2015) is a method of dynamic selection of committees that uses the idea of selection by meta-learning. In it, a meta-problem is created to determine whether a classifier is competent or not for a given test instance. According to (DANTAS, 2021) the META-DES uses a total of five different criteria for the extraction of meta-features, to establish the new region of the metaproblem.

After that, a metaclassifier is trained, based on the meta-features to identify whether a classifier is competent or not to classify a sample. Raters marked as competent will be chosen to compose the test instance's rating committee. Figure 3 presents the META-DES committee selection scheme, divided into two phases.

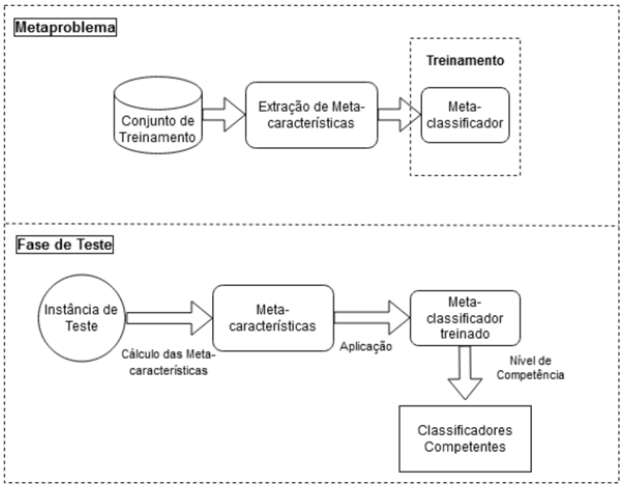

Fig 3. Scheme of operation of META-DES (DANTAS, 2021)

In the first phase, the metaproblem is defined. In it, a set of metafeatures is extracted from the training set. Soon after, a meta-classifier is trained based on the extracted meta-features. This meta-classifier will be used to determine the competence level of the base classifiers for each test instance.

In the second phase, a test instance is presented to the selection scheme, its meta-features are calculated and presented to the previously trained meta-classifier. The meta-classifier then estimates the confidence level for each base classifier. Classifiers judged competent will be selected to compose the committee and carry out the classification process of the test instance.

III. DYNAMIC FUSION

The approach proposed in this research follows some steps. Initially, a test was applied with combinations of Pool and K Region of competence. A Pool of 5, 10, 15, 20, 25 and 30 Decision Trees, generated through the Bagging method, was applied. In the KNORA-E and META-DES methods, the K value of 3, 7 and 11 were applied to define the region of competence. The value of K also served for the new dynamic combination method.

In the Dynamic Fusion method proposed in this work, the same value of K was applied to the KNORA-E and META-DES methods, when the competence region is traced, the basic method that is most correct in that region will be chosen to classify the new sample. In case of a tie, the value of K in the Dynamic Fusion is incremented until there is a tie between the methods that had a tie. The ten merger methods used are: Majority Vote; Average; Maximum; Minimum; Geometric Mean; Naive Bayes; Edge and Three types of Artificial Neural Networks.

In the Artificial Neural Network (HARD) the classes resulting from the outputs of the base classifiers selected as input (X features) are used for a Multi-layer Perceptron Neural Network to be trained/tested and, finally, to make the final decision of the chosen class. In unselected classifiers the output in the sample is used -1.

In the other two Artificial Neural Networks, the percentages of predictions of each class resulting from the outputs of the selected base classifiers are used. In one (SOFT) classifiers not selected the output in the sample is used -1 and in the other (SOFT CLASS) 1/ number of total classes of the base is used.

Two types of weighted merger were also applied, one by weighted weight, and another by weighted vote. But these two methods were applied only for comparative purposes and were not used in the proposed dynamic fusion method.

When combining by Weighted on the probabilities of the selected classifiers, the weight is 1 / distance of classes. In the weighted vote, the weight in the votes of the selected classifiers of 1 / distance of classes was used.

Thus, twelve fusion methods were applied: Majority Vote, Average, Maximum, Minimum, Geometric mean, Neural Network HARD, Neural Network SOFT, Neural Network SOFT CLASS, Edge, Naive Bayes, Weighted committee weighting and Weighted committee voting. The method proposed in this work (dynamic fusion) chooses for each new sample which of the ten fusion methods (Majority Vote, Average, Maximum, Minimum, Geometric mean, Neural Network HARD, Neural Network SOFT, Neural Network SOFT class, Edge, Naive Bayes) use. Finally, there are thirteen results that will be compared with each other.

The databases used in this work are available in the *UCI Machine Learning Repository*[1], in table 1 some characteristics of the databases are presented.

TABLE I. DATABASE CHARACTERISTICS

Name	Examples	Attributes	Class
Cardiac insufficiency	368	53	2
Car	1728	6	4
Seismic-bumps	2584	18	2
Zoo	101	16	7
Ionosphere	351	34	2
Prognostic	198	33	2
Wine	178	13	3
Dermatology	366	34	6
Heatr	303	13	2
Bone marrow	187	36	2
Algerian Forest Fires	244	13	2
Congressional Voting Records	435	16	2
Maternal Health Risk	1014	6	3
Risk Factors Cervical Cancer	855	28	2
Phishing Website	2456	30	2

Each database is divided into a proportion of 50% training, 16.7% Testing 16.7% validation 1, 16.6% validation 2. The training database was used to generate the initial classifiers and the committee models. The test set was used to score each type of fusion, and the dynamic fusion. The Validation 1 base was used to train the methods of Artificial Neural Networks and Naive Bayes. Validation set 2 was used to obtain the competence region in the proposed dynamic fusion method. Each test was applied 30 times the average obtained. The results for each base for each type of fusion using the combination of Pool and Region of competence K were 18. Totaling a total of 270 results per technique overall. These results were later used in statistical tests. Tables 2 and 3 show the averages of the 18 Scores for each type of classifier fusion using the KNORA-E and META-DES technique.

The results presented initially show an improvement in the overall average through Dynamic Fusion in relation to other mergers. In KNORA-E, with the proposed method it was possible to obtain improvements in the general result of 4 of the 15 databases. Some results show improvements of more than 1 percentage point on average. Which can represent even better results in combinations of pool and region of competence.

In META-DES, however, there were better results in only 2 of the 15 databases. Again, showing improvements of more than 1 percentage point in the Maternal Health Risk baseline. The Neural Network SOFT class presented the best results among the Networks, and in some bases, it also stood out.

[1] https://archive.ics.uci.edu/ml/index.php

TABLE II. SCORE OF FUSION METHODS IN KNORA-E

x	Majority Vote	AVG	Maximum	Minimum	Geometric average	*Weighted weight in committee*	*Weighted voting in committee.*	*Neural Network HARD*	*Neural Network SOFT*	*Neural Network SOFT CLASS*	*Edge*	*Naive Bayes*	*Dynamic Fusion K*
Cardiac insufficiency	94.58	94.58	95.17	95.17	95.17	88.77	88.77	95.05	94.17	95.02	94.58	92.93	95.66
Car	97.30	97.30	93.35	91.81	91.81	84.90	84.90	94.29	96.12	97.15	96.39	93.85	97.03
Seismic-bumps	92.07	92.06	93.06	93.06	93.06	91.63	91.64	92.39	91.70	92.11	92.07	85.43	92.57
Zoo	93.63	93.63	93.10	89.97	89.97	76.47	76.47	90.59	88.37	92.45	90.82	76.21	93.73
Ionosphere	90.97	90.97	79.84	79.84	79.84	77.02	77.02	88.21	89.20	88.96	90.97	84.31	89.91
Prognostic	69.48	69.48	44.75	44.75	44.75	62.09	62.09	69.53	68.86	69.28	69.48	71.80	71.62
Wine	93.68	93.68	78.35	72.89	72.89	64.27	64.27	85.34	91.25	92.99	92.61	57.16	93.10
Dermatology	95.84	95.84	90.94	84.75	84.75	78.84	78.84	90.79	93.79	95.82	94.03	78.01	95.74
Heatr	76.90	76.90	63.44	63.44	63.44	67.71	67.71	74.41	73.68	72.61	76.90	73.01	76.04
Bone marrow	91.70	91.70	88.39	88.39	88.39	90.97	90.97	90.13	88.48	89.78	91.70	89.18	91.45
Algerian Forest Fires	97.65	97.65	96.13	96.13	96.13	97.28	97.28	96.57	96.72	96.90	97.65	96.88	97.63
Congressional Voting Records	94.45	94.45	91.70	91.70	91.70	92.46	92.46	93.11	93.51	93.36	94.45	92.80	93.94
Maternal Health Risk	77.48	77.45	78.12	77.98	77.98	64.80	64.80	75.61	74.56	75.73	77.62	73.40	78.69
Risk Factors Cervical Cancer	91.47	91.47	93.21	93.21	93.21	90.97	91.00	92.57	92.17	92.36	91.47	91.05	93.17
Phishing Website	95.48	95.48	92.48	92.48	92.48	92.50	92.50	94.85	94.92	94.77	95.48	90.99	95.49
Overall Average	**90.18**	**90.18**	**84.80**	**83.70**	**83.70**	**81.38**	**81.38**	**88.23**	**88.50**	**89.29**	**89.75**	**83.13**	**90.38**

TABLE III. SCORE OF FUSION METHODS IN META-DES

x	Majority Vote	AVG	Maximum	Minimum	Geometric average	*Weighted weight in committee*	*Weighted voting in committee.*	*Neural Network HARD*	*Neural Network SOFT*	*Neural Network SOFT CLASS*	*Edge*	*Naive Bayes*	*Dynamic Fusion K*
Cardiac insufficiency	94.23	94.23	94.52	94.52	94.52	90.43	90.43	94.07	95.09	95.22	94.23	91.30	95.01
Car	96.80	96.80	89.89	87.12	87.12	83.86	83.86	93.72	96.77	96.99	95.68	91.08	96.85
Seismic-bumps	92.01	92.01	93.25	93.25	93.25	92.37	92.37	92.37	92.17	92.62	92.01	82.59	92.66
Zoo	94.54	94.54	89.54	85.26	85.26	75.92	75.92	90.88	95.13	95.26	89.80	75.13	94.48
Ionosphere	90.46	90.46	80.22	80.22	80.22	76.99	76.99	88.31	89.42	88.96	90.46	83.57	89.67
Prognostic	74.26	74.26	41.30	41.30	41.30	65.80	65.80	69.21	69.65	70.28	74.26	58.61	73.16
Wine	92.84	92.84	74.37	67.87	67.87	64.39	64.39	84.39	91.97	92.45	91.97	58.95	92.36
dermatology	96.22	96.22	88.23	80.62	80.62	78.96	78.96	90.49	96.10	96.17	93.85	78.53	96.32
Heatr	78.54	78.54	63.02	63.02	63.02	68.59	68.59	74.18	75.70	76.46	78.54	66.94	77.85
Bone marrow	92.01	92.01	89.43	89.43	89.43	91.49	91.49	89.77	90.68	90.82	92.01	88.87	91.79
Algerian Forest Fires	97.75	97.75	95.86	95.86	95.86	97.42	97.42	96.58	96.82	96.78	97.75	96.54	97.71
Congressional Voting Records	94.54	94.54	92.17	92.17	92.17	92.72	92.72	93.09	93.70	93.50	94.54	92.72	94.08
Maternal Health Risk	77.54	77.62	77.10	76.29	76.28	64.49	64.51	75.55	76.67	77.32	77.74	68.05	78.93
Risk Factors Cervical Cancer	91.85	91.84	93.66	93.66	93.66	92.16	92.19	92.54	92.19	92.32	91.85	89.02	93.40
Phishing Website	95.50	95.51	92.17	92.17	92.17	92.63	92.62	94.92	95.18	95.12	95.50	90.36	95.40
Overall Average	**90.61**	**90.61**	**83.65**	**82.18**	**82.18**	**81.88**	**81.88**	**88.01**	**89.82**	**90.02**	**90.01**	**80.82**	**90.64**

IV. STATISTICAL ANALYSIS

To confirm that these data actually show better results, the Friedman test (FRIEDMAN, 1940) was applied to verify if there are statistical differences between the 13 fusions of classifiers. Each technique has 270 results, obtained from combinations of Pool and Region of Competence. The Friedman test is used to compare linked sample data. It is used to be able to state the hypothesis that the k related observations derive from the same population.

The significance level used was *p = 0.05*. If the p value is less than the established value, the null hypothesis is rejected, guaranteeing a confidence level greater than 95%. Tables IV and V show the results of the Friedman test.

TABLE IV. FRIEDMAN TEST IN KNORA-E

Chi-Squared	**df**	**P**	**Kendall's W**
1.161.254	12	< 0.05	0.358

TABLE V. FRIEDMAN TEST IN META-DES

Chi-Squared	**df**	**P**	**Kendall's W**
960.454	12	< 0.05	0.296

As the result shows that there are statistical differences, the Critical Difference Diagram (DEMSAR, 2006) is used to verify what these differences are.

In the Critical Difference Diagram, the comparison between pairs is statistically significant if the difference between the average of the rankings between them is greater than the value obtained by calculating the Critical Difference Diagram (CD).

Figure 4 shows the fusion methods CD in KNORA-E. Those methods that have a greater distance than the line representing the CD value are considered statistically different.

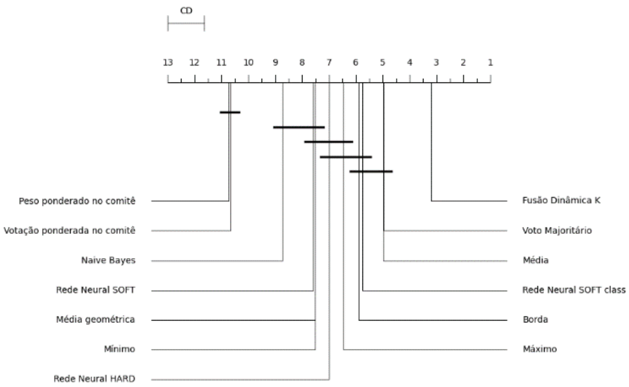

Fig 4. Critical Diference Diagram in KNORA-E

The dynamic fusion method presents statistical differences to all others. Fusion by average, Majority Vote, Edge and Neural Network SOFT CLASS present good results, but below the proposed in this work. So, the CD shows that in fact in KNORA-E dynamic fusion improves the results statistically.

Figure 5 shows the CD of mergers of classifiers in META-DES. In this case, in spite of Dynamic Fusion having the best results, this cannot be stated statistically. Because the CD did not find statistical differences between dynamic fusion, mean, majority vote and soft class neural network.

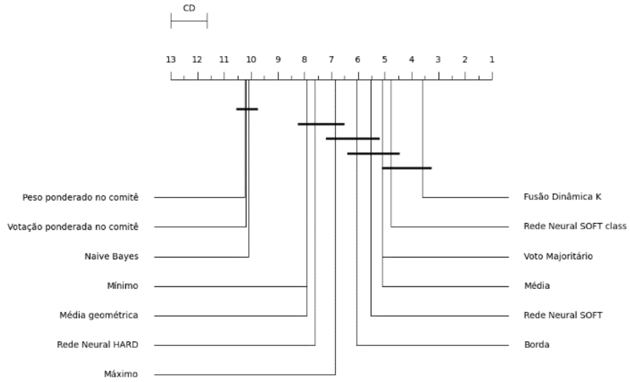

Fig 5. Critical Diference Diagram in META-DES

In both cases techniques such as weighted weighting and weighted voting showed statistically equal results, similarly the maximum, minimum and geometric mean as well.

V. CONCLUSION

In this research, a dynamic value fusion technique was proposed that uses the competence region to verify which fusion technique is best to use for a new sample and a given data region.

The technique uses 10 possible mergers of classifiers: Majority Vote, Average, Maximum, Minimum, Geometric Average, Edge, Naive Bayes and three types of Neural Network. Fusion techniques were applied through KNORA-E and META-DES.

Classifier pool configurations of 5, 10, 15, 20, 25, 30 and competence region of 3,7,11 were used. Thus, the type of merger selected was the one that best matched the most similar neighbors.

The results showed a better average score in relation to the other fusion methods, both in KNORA-E and in META-DES. Despite this, only in KNORA-E were statistical differences between the proposed method and the others presented. Because there are only 270 results, it may have influenced the fact that there was no statistical difference in the proposed method and the others in the META-DES.

In general, the proposed technique showed good results and indeed improvements. Thus, contributing to the classifiers committee area. Because combinations were made with several methods, it may have interfered a little. Studies are needed to better verify which fusion methods can be used dynamically. Because removing some types of fusion can expand the results.

This research was limited to the fact that only 15 databases were used and with few classes. It is necessary to expand the tests and check individually in each database whether or not there are improvements in dynamic fusion.

In future studies these limitations will be overcome, increasing the number of databases, applying statistical tests individually in each database and expanding the tests for KNORA-Union and Overrall Local Accuraccy (OLA).

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].