TWO-STEP SUPERVISED CONFIDENCE MEASURE FOR AUTOMATIC FACE RECOGNITION

Ladislav Lenc[1,2], *Pavel Král*[1,2]

[1]Dept. of Computer Science & Engineering
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic

[2]New Technologies for the Information Society
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic

ABSTRACT

This paper deals with automatic face recognition in the context of a real application for the Czech News Agency. This system will be used to annotate people in photographs during insertion into the database. Unfortunately, the accuracy of the current face recognition approaches is limited and therefore another task to process the recognition results is very important. The main contribution of this work thus consists in proposing and evaluating a novel supervised confidence measure method as the post-processing step in order to detect incorrectly classified face images from the classifier's output. We experimentally show that the proposed confidence measure is beneficial for our application.

***Index Terms*—** Face Recognition, Czech News Agency, Confidence Measure, Multi-layer Perceptron, Scale Invariant Feature Transform (SIFT)

1. INTRODUCTION

Automatic Face Recognition (AFR) consists in identification of a person from an image or from a video frame by a computer. This field has been intensively studied by many researchers during the past few decades and nowadays, it can be seen as one of the most progressive biometric authentication methods. Numerous AFR methods have been proposed and the face recognition has become the key task in several applications as for instance surveillance of wanted persons, access control to restricted areas, automatic annotation of the photos used in the recently very popular photo sharing applications or in the social networks, and so on.

The most of the proposed approaches perform well in the laboratory conditions where the images are well aligned, the face pose and lighting conditions are similar, etc. Unfortunately, their performance is significantly decreased when these conditions are not accomplished. Several methods have been introduced to handle these limitations, but none of them performs well in a fully uncontrolled environment.

In our previous work, we proposed the SIFT based Kepenekci face recognition method [1]. We showed that this method significantly outperforms the other approaches particularly on lower quality real data. However, the face recognition rates are still far to be perfect.

The main goal of this paper thus consists in proposing a novel Confidence Measure (CM) technique in order to detect and handle incorrectly recognized samples. The proposed CM has two steps. The first unsupervised step is based on the *posterior* class probability, while the second step, the supervised one, uses a multi-layer perceptron as a classifier. Note that to the best of our knowledge, there is no similar two-step CM approach available. The other known approaches are composed of only one step.

The results of this work will be used by the Czech News Agency (ČTK[1]) to annotate people in photographs during insertion into the photo-database[2]. The main issue is to annotate only the correctly identified persons. The incorrectly recognized faces must be detected and their face labels manually assigned.

The paper structure is as follows. The following section gives a brief overview of important face recognition and confidence measure methods. Section 3 describes our AFR method. This section also details the proposed confidence measure approach. Section 4 evaluates and compares the performance of our confidence measure on the ČTK corpus. In the last section we discuss the achieved results and give some further research directions.

2. RELATED WORK

This section is composed of two parts. The successful face recognition approaches are described in the first part, while the second part is focused on the confidence measure task itself.

[1]http://www.ctk.eu
[2]http://multimedia.ctk.cz/en/foto/

2.1. Face Recognition

One of the first successful approaches is Eigenfaces [2]. This approach is based on the Principal Component Analysis (PCA). Unfortunately, it is sensitive on variations in lighting conditions, pose and scale. However, the PCA based approaches are still popular, as shown in [3].

Another method, the Fisherfaces [4], is derived from Fisher's Linear Discriminant (FLD). Similarly to Eigenfaces, it projects images into less dimensional space. According to the authors, this approach should be less sensitive to changing lighting conditions than Eigenfaces. A recent extension of this approach, called L-Fisherfaces, is proposed in [5].

Independent Component Analysis (ICA) can be also successfully used in the automatic face recognition field [6]. Contrary to Eigenfaces, ICA uses higher order statistics. It thus provides more powerful data representation. Bartlett et al. showed in [7] that ICA performs slightly better than PCA method on the FERET [8] corpus.

Authors of [9] propose another efficient AFR approach called Locality Pursuit (LP). This method uses locality preserving projections in the high-dimensional whitened space. Experimental results show that the LP approach achieves higher accuracy than ICA on the FERET corpus.

Another efficient AFR approach is the Elastic Bunch Graph Matching (EBGM) [10]. This approach uses features constructed by the Gabor wavelet transform. Several other successful approaches based on Gabor wavelets have been introduced [11]. Some approaches [12] combine the pre-processing with Gabor wavelets with well-established methods such as Eigenfaces, Fisherfaces, etc.

Kepenekci proposes in [13] an algorithm that addresses the main issue of Elastic Bunch Graph Matching, manual labelling of the landmarks. The proposed method outperforms the classical EBGM.

Speeded-Up Robust Features (SURF) [14] is another recent method used for automatic face recognition. This method is invariant to face rotation. To ensure rotation invariance, one orientation is assigned to each key-point. The computation is based on the circular neighbourhood of the key-points.

Recently, the Scale Invariant Feature Transform (SIFT) is successfully used for face recognition [15]. The main advantage of this approach is the ability to detect and describe local features in images. The features (see Figure 1) are invariant to image scaling, translation and rotation. Moreover, they are also partly invariant to changes in illumination. Therefore, this approach is beneficial for face recognition in real conditions where the images differ significantly. Another approach based on the SIFT, called Fixed-key-point-SIFT (FSIFT), is presented in [16].

For further information about the face recognition, please refer to the surveys [17, 18].

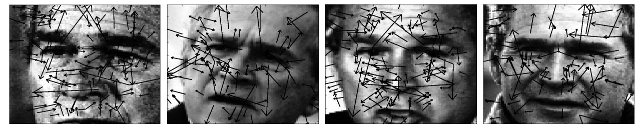

Fig. 1. SIFT features depicted in a face image

2.2. Confidence Measure

Confidence measure is used as a post-processing of the recognition to determine whether a result is correct or not. The incorrectly recognized samples should be removed from the recognition set or another processing (e.g. manual correction) can be further realized. This technique is mainly used in the automatic speech processing field [19, 20, 21, 22] and is mostly based on the *posterior* class probability. However, it can be successfully used in another research areas as shown in [23] for genome maps construction, in [24] for stereo vision or in [25] for handwriting sentence recognition.

Another approach related to the confidence measure is proposed by Proedrou et al. in the pattern recognition task [26]. The authors use a classifier based on the nearest neighbours algorithm. Their confidence measure is based on the algorithmic theory of randomness and on transductive learning.

Unfortunately, only few works about the confidence measure in the face recognition domain exist. Li and Wechsler propose a face recognition system which integrates a confidence measure [27] in order to reject unknown individuals or to detect incorrectly recognized faces. Their confidence measure is, as in the previous case, based on the theory of randomness. The proposed approaches are validated on the FERET database.

Eickeler et al. propose and evaluate in [28] five other CMs also in the face recognition task. They use a pseudo 2-D Hidden Markov Model classifier with features created by the Discrete Cosine Transform (DCT). Three proposed confidence measures are based on the *posterior* probabilities and two others on ranking of results. Authors experimentally show that the *posterior* class probability gives better results for the recognition error detection task.

3. FACE RECOGNITION WITH CONFIDENCE MEASURE

3.1. Face Recognition

For the face recognition task, we use our previously proposed SIFT based Kepenekci method [1] which uses an efficient SIFT algorithm for parametrization and adapted Kepenekci matching [29] for recognition. This method was chosen, because as proven previously, it significantly outperforms the other approaches particularly on lower quality real data.

3.1.1. *SIFT Parametrization*

This algorithm creates an image pyramid with re-sampling between each level to determine potential key-point positions. Each pixel is compared with its neighbours. Neighbours in its level as well as in the two neighbouring levels are analysed. If the pixel is maximum or minimum of all neighbouring pixels, it is considered to be a potential key-point.

For the resulting set of key-points their stability is determined. The locations with low contrast and unstable locations along edges are deleted.

The orientation of each key-point is computed next. The computation is based on gradient orientations in the neighbourhood of the pixel. The values are weighted by the magnitudes of the gradient.

The last step consists in the descriptor creation. The computation involves the 16×16 neighbourhood of the pixel. Gradient magnitudes and orientations are computed in each point of the neighbourhood. Their values are weighted by a Gaussian curve. For each sub-region of size 4×4 (16 regions), the orientation histograms are created. Finally, a vector containing 128 (16×8) values is created.

3.1.2. *Adapted Kepenekci Matching*

This approach combines two methods of matching and uses the weighted sum of the two results.

Let T be a test image and G a gallery image. For each feature vector t of face T we determine a set of relevant vectors g of face G. Vector g is relevant iff:

$$\sqrt{(x_t - x_g)^2 + (y_t - y_g)^2} < distanceThreshold \quad (1)$$

where x and y are the coordinates of the feature vector points.

If no relevant vector to vector t is identified, vector t is excluded from the comparison procedure. The overall similarity of two faces OS is computed as the average of similarities S between each pair of corresponding vectors as:

$$OS_{T,G} = mean\,\{S(t,g), t \in T, g \in G\} \quad (2)$$

Then, the face with the most similar vector to each of the test face vectors is determined. The C_i value denotes how many times gallery face G_i was the closest one to some of the vectors of test face T. The similarity is computed as C_i/N_i where N_i is the total number of feature vectors in G_i. The weighted sum of these two similarities FS is used for similarity measure:

$$FS_{T,G} = \alpha OS_{T,G} + (1-\alpha)\frac{C_G}{N_G} \quad (3)$$

The optimal value of the parameter α is set experimentally on the development corpus.

The face is recognized by the following equation:

$$\hat{FS_{T,G}} = \arg\max_G (FS_{T,G}) \quad (4)$$

The cosine similarity is used for vector comparison.

3.2. Confidence Measure

The confidence measure is used after the recognition itself in order to identify and remove incorrectly recognized faces from the resulting set. The proposed approach is composed of the two steps which are presented next.

3.2.1. *Step - I.*

As in many other papers [21, 30, 22], this step is based on the estimation of the *posterior* class probability.

Let the output of the classifier be $P(F|C)$, where C is the recognized face class and F represents the face features. The values $P(F|C)$ are normalized to compute the *posterior* class probabilities as follows:

$$P(C|F) = \frac{P(F|C).P(C)}{\sum_{I \in \mathcal{FIM}} P(F|I).P(I)} \quad (5)$$

$\mathcal{FIM}$ represents the set of all individuals and $P(C)$ denotes the *prior* probability of the individual's (face) class C.

We propose two different approaches for this task. In the first approach, called *absolute confidence value*, only faces $\hat{C}$ complying with

$$\hat{C} = \arg\max_C (P(C|F)) \quad (6)$$

$$P(\hat{C}|F) > T \quad (7)$$

are considered as being recognized correctly.

The second approach, called *relative confidence value*, computes the difference between the *best* score and the *second best* one by the following equation:

$$P\Delta = P(\hat{C}|F) - \max_{C \neq \hat{C}} (P(C|F)) \quad (8)$$

Only the faces with $P\Delta > T$ are accepted. This approach aims to identify the "dominant" faces among all the other candidates. T is the acceptance threshold and its optimal value is adjusted experimentally.

The confidence measure values of these two approaches will be hereafter represented by the variable R.

These two approaches should work separately as already presented in [31]. However, we suppose that the next supervised step will significantly improve the results.

3.2.2. Step - II.

In this step, we use the score R obtained in the previous step I. as an input. We use a Multi-layer Perceptron (MLP) to model *posterior* probability $P(H|R)$. The variable H has only two values and determines whether the face image was classified correctly or not.

Three MLP configurations are built and evaluated:

1. supervised *absolute confidence value* method,

2. supervised *relative confidence value* method,

3. combination of methods *1* and *2*

The MLP topology will be described in detail in the experimental section. Note that this second step use only the score obtained in the first step and is thus completely independent of the T value.

4. EXPERIMENTS

4.1. Czech News Agency Corpus

This corpus is composed of images of individuals in an uncontrolled environment that were randomly selected from the large ČTK database. All images were taken over a long time period (20 years or more). The corpus contains gray-scale images of 638 individuals of size 128×128 pixels. It contains about 10 images for each person. The orientation, lighting conditions and image backgrounds differ significantly.

Figure 2 shows examples of one face from this corpus. This corpus is available for free for research purposes at `http://home.zcu.cz/~pkral/sw/` or upon request to the authors.

Fig. 2. Examples of one face from the ČTK face corpus

4.2. Recognition Results with Confidence Measure

In the following experiments, we recognize 638 individuals. We use one example/person for testing and the remaining samples for training. The accuracy of the confidence measure approaches is evaluated on the testing examples.

4.2.1. Accuracy of the first CM Step

In the first experiment we would like to show the performance of the separate first step of our confidence measure. As in many other articles in the confidence measure field, we will use the Receiver Operating Characteristic (ROC) curve [32] for evaluation. This curve clearly shows the relationship between the true positive and the false positive rate for the different *acceptance* threshold.

Figure 3 shows the results of the *absolute confidence value* method, while the results of the *relative confidence value* approach are given in Figure 4. These figures show that both approaches are suitable for our task in order to identify incorrectly recognized faces. Moreover, the *relative confidence value* method significantly outperforms the *absolute confidence value* approach. Better accuracy of this approach can be explained by the fact that the significantly higher *posterior* probability (among all the other candidates) is a better metrics than the simple absolute value of this probability.

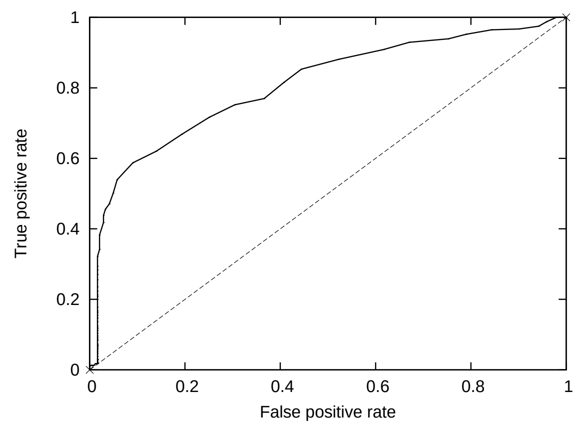

Fig. 3. ROC curve of the *absolute confidence value* method

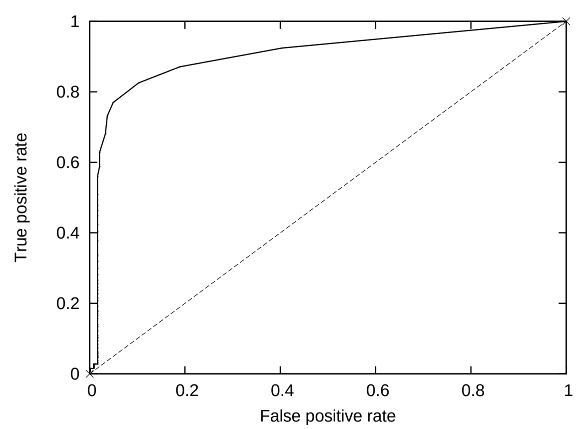

Fig. 4. ROC curve of the *relative confidence value* method

The first section of the Table 1 shows the scores of the optimal threshold configurations. The F-measure (F-mes) [33] is used as an evaluation metric, the Precision (Prec) and Recall (Rec) are also reported in this table. These two values have similar importance for our application. Therefore, the optimal threshold value $\hat{T}$ has been defined as the "best" compromise between these two values as follows:

$$\hat{T} = \arg\min_{T} \mid 1 - \frac{Prec}{Rec} \mid \tag{9}$$

The second remark is that the results of the first step of the proposed CM approach are used as our baseline, because it was successfully used in our previous work [31] and in some other approaches [21, 30] as well.

4.2.2. *Accuracy of the whole CM approach*

In the last experiment, we would like to evaluate the results of the whole proposed confidence measure method, i.e. after the second step with an MLP. The best MLP topology uses three layers (in all cases): one or two input neurons, 10 neurons in the hidden layer and two outputs (*correctly* and *incorrectly* recognized face). One input is used in the case when the supervised CM methods are used separately and two input neurons are used when we combine both approaches. The MLP topology was defined empirically on a small development corpus which contains 120 examples (i.e. 120 confidence values). Note, that this corpus has been created fully automatically by the *relative confidence value* method.

The results of this experiment are reported in the second section of the Table 1. These results clearly show that the second supervised step of the proposed confidence measure is very important. The F-measure improvement is about 23% in absolute value over the baseline approach. This large improvement may be explained by a generally better performance of the supervised methods (*step II.* in comparison with the unsupervised ones (*step I.*). This table further shows that the second, supervised *relative confidence value* method, significantly outperforms the first, unsupervised *absolute confidence value*, approach. However, the first method brings further relevant information regarding the second approach. This fact is confirmed by the result of the combined approach, which gives the best recognition score from the whole.

No	Confidence Measure	Prec	Rec	F-mes
1. Performance after the first CM step				
1.	*absolute confidence value* method	65.7	60.6	63.0
2.	*relative confidence value* method	69.6	60.8	64.9
2. Performance of the whole confidence measure approach				
3.	supervised *absolute confidence value* method	89.8	68.1	77.5
4.	supervised *relative confidence value* method	94.1	82.5	87.9
5.	combination of methods *3* and *4*	95.6	87.9	91.6

Table 1. Performance of the confidence measure approaches [in %]

5. CONCLUSIONS AND PERSPECTIVES

In this work, we proposed and evaluated the new confidence measure approach in the automatic face recognition task. This technique is used in order to detect and handle incorrectly recognized samples. The proposed approach is composed of two steps: the first step is based on the *posterior* class probability, while the second step uses an MLP classifier. The results show that the second step of the proposed approach is very important. The F-measure improvement is about 23% in absolute value over the baseline approach. Therefore, we conclude that the proposed confidence measure will be integrated into our application for the ČTK.

The first perspective consists in proposing other confidence measures in the post-processing step. These methods will be based on the main properties of the face model. We further assume combining them with the confidence measure proposed in this paper. Another future work can consist in the progressive adaptation of the confidence measure model depending on the recognized data.

6. ACKNOWLEDGEMENTS

This work has been partly supported by the UWB grant SGS-2013-029 Advanced Computer and Information Systems and by the European Regional Development Fund (ERDF), project "NTIS - New Technologies for Information Society", European Centre of Excellence, CZ.1.05/1.1.00/02.0090. We also would like to thank Czech New Agency (ČTK) for support and for providing the photographic data.

7. REFERENCES

[1] L. Lenc and P. Král, "Face recognition under real-world conditions," in *International Conference on Agents and Artificial Intelligence*, Barcelona, Spain, February 14-18 2013.

[2] Matthew A. Turk and Alex P. Pentland, "Face recognition using eigenfaces," in *IEEE Computer Society Conference on In Computer Vision and Pattern Recognition*. Computer Vision and Pattern Recognition, 1991.

[3] Bruce Poon, M Ashraful Amin, and Hong Yan, "Performance evaluation and comparison of pca based human face recognition methods for distorted images," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 4, pp. 245–259, 2011.

[4] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997.

[5] Cheng-Yuan Zhang and Qiu-Qi Ruan, "Face recognition using l-fisherfaces," *Journal of information science and engineering*, vol. 26, no. 4, pp. 1525–1537, 2010.

[6] Bruce A. Draper, Kyungim Baek, Marian Stewart Bartlett, and J. Ross Beveridge, "Recognizing faces with PCA and ICA," in *Computer Vision and Image Understanding*, 2003, pp. 115–137.

[7] Marian Stewart Bartlett, Javier R. Movellan, and Terrence J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, pp. 1450–1464, 2002.

[8] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.

[9] W. Deng, Y. Liu, J. Hu, and J. Guo, "The small sample size problem of ICA: A comparative study and analysis," *Pattern Recognition*, vol. 45, no. 12, pp. 4438–4450, 2012.

[10] David S. Bolme, *Elastic Bunch Graph Matching*, Ph.D. thesis, Colorado State University, 2003.

[11] L. Shen and L. Bai, "A review on gabor wavelets for face recognition," *Pattern Analysis & Applications*, 2006.

[12] L. Shen, *Recognizing Faces - An Approach Based on Gabor Wavelets*, Ph.D. thesis, University of Nottingham, 2005.

[13] B. Kepenekci, *Face Recognition Using Gabor Wavelet Transform*, Ph.D. thesis, The Middle East Technical University, 2001.

[14] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, June 2008.

[15] M. Aly, "Face recognition using sift features," 2006.

[16] J. Krizaj, V. Struc, and N. Pavesic, "Adaptation of sift features for robust face recognition," 2010.

[17] Paola Campadelli and Raffaella Lanzarotti, "A face recognition system based on local feature characterization," in *Advanced Studies in Biometrics*, pp. 147–152. Springer, 2005.

[18] M. P. Beham and S. M. M. Roomi, "A review of face recognition methods," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 4, 2013.

[19] Grégory Senay, Georges Linares, and Benjamin Lecouteux, "A segment-level confidence measure for spoken document retrieval," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5548–5551.

[20] Grégory Senay and Georges Linares, "Confidence measure for speech indexing based on latent dirichlet allocation.," in *INTERSPEECH*, 2012.

[21] H. Jiang, "Confidence measures for speech recognition: A survey," *Speech Communication*, vol. 45, no. 4, pp. 455–470, 2005.

[22] Frank Wessel, Ralf Schluter, Klaus Macherey, and Hermann Ney, "Confidence measures for large vocabulary continuous speech recognition," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 3, pp. 288–298, 2001.

[23] Bertrand Servin, Simon de Givry, and Thomas Faraut, "Statistical confidence measures for genome maps: application to the validation of genome assemblies," *Bioinformatics*, vol. 26, no. 24, pp. 3035–3042, 2010.

[24] X. Hu and P. Mordohai, "A quantitative evaluation of confidence measures for stereo vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2121–2133, 2012.

[25] Sanparith Marukatat, Thierry Artières, Patrick Gallinari, and Bernadette Dorizzi, "Rejection measures for handwriting sentence recognition," in *Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on*. IEEE, 2002, pp. 24–29.

[26] Kostas Proedrou, Ilia Nouretdinov, Volodya Vovk, and Alexander Gammerman, "Transductive confidence machines for pattern recognition," in *ECML'02*, 2002, pp. 381–390.

[27] Fayin Li and Harry Wechsler, "Open world face recognition with credibility and confidence measures," in *Audio-and Video-Based Biometric Person Authentication*. Springer, 2003, pp. 462–469.

[28] Stefan Eickeler, Mirco Jabs, and Gerhard Rigoll, "Comparison of confidence measures for face recognition.," in *FG*. 2000, pp. 257–263, IEEE Computer Society.

[29] L. Lenc and P. Král, "Novel matching methods for automatic face recognition using sift," in *8th AIAI (Artificial Intelligence Applications and Innovations) Confence*, Halkidiki, Greece, September 27-30 2012.

[30] Stefan Eickeler, Mirco Jabs, and Gerhard Rigoll, "Comparison of confidence measures for face recognition," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 257–262.

[31] L. Lenc and P. Král, "Confidence measure for automatic face recognition," in *International Conference on Knowledge Discovery and Information Retrieval*, Paris, France, October 26-29 2011.

[32] Christopher D Brown and Herbert T Davis, "Receiver operating characteristics curves and related decision measures: A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 80, no. 1, pp. 24–38, 2006.

[33] DMW Powers, "Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.