ISSN 1870-4069

A Copula Functions Approach for Predicting Temporomandibular Disorders

Carlos López-Hernández[1], Rogelio Salinas-Gutiérrez[1], Omar Gutiérrez-Navarro[2], Ivonne Bazán-Trujillo[2]

[1] Universidad Autónoma de Aguascalientes, Centro de Ciencias Básicas, Mexico

[2] Universidad Autónoma de Aguascalientes, Centro de Ciencias de la Ingeniería, Aguascalientes, Mexico

{al223348, rogelio.salinas, omar.gutierrezn, ivonne.bazan }@edu.uaa.mx

Abstract. Temporomandibular disorders (TMD) are a set of conditions affecting the temporomandibular joint and associated muscles. This work addresses the classification of healthy and TMD diagnosed individuals using supervised learning models. 118 features were extracted from mandibular movement signals and a selection of those features is used for training the classification models. The dataset of signals was obtained via a marker tracking system with a depth camera. A set of 5 classifiers including two based on copula functions that model dependence structures is tested. The performances of these classifiers are compared. Results show good performance for the copula classifiers, although they do not show better performance than those of models such as KNN, SVM or Naive Bayes.

Keywords: Temporomandibular disorders, feature dependencies, supervised classification, copula functions.

1 Introduction

Temporomandibular disorders (TMD) are a group of musculoskeletal conditions that involve the temporomandibular joint, the masticatory muscles and all associated tissues. TMD are one of the most common causes of pain in the mouth and face, and have the potential to produce chronic pain [6]. The causes behind TMD are multifactorial, including biological, psychological and social causes. The diagnosis of TMD depends largely on clinical history and physical evaluations such as the DC-TMD of [13].

There are several works addressing the classification of TMD using supervised learning models. They use different kind of data such as sound [15], movement [2], or EMG [16]. With this data, models such as KNN, naive Bayes, or SVM have been trained. However, to the best of our knowledge there has not been attempts

of predicting TMD with copula function models. Copula functions are joint distribution functions that provide a flexible way of modeling association among features. They enable to separate the dependence structure from the marginal distributions of the features[11]. Being able to model both linear and non linear dependencies, copula functions have been used in fields such as economics and finance [5], hydrology [8], evolutionary computation [12], multispectral image processing [17] among others.

This work proposes the use of Frank and Gaussian bivariate copulas for modeling dependencies among features and a tree graphical model for the selection of the strongest association. Probabilistic classifiers based on these copulas are trained in the classification of healthy individuals and those with temporomandibular disorders, aiming for an improvement performance based on the dependency information provided by the copula functions. A Mandibular movement dataset from healthy and TMD diagnosed individuals is used for training the models. This dataset was obtained via a marker tracking system with a depth camera.

The content of the paper is as follows: Section 2 provides a brief introduction to the theory of copula functions with emphasis on the selected copulas and the probabilistic classifier. Section 3 describes data used in this article; Section 4 narrates the methodological process followed. The results of the experiments are found in Section 5, and finally Section 6 concludes the paper summarizing the work done and our findings.

2 Copula Theory

2.1 Copula Functions

Copula functions were first introduced by Sklar [14] as a way to differentiate the effect of dependencies from that of marginal distributions in a joint distribution function. By using copula functions it is possible to model both linear and nonlinear dependencies among features.

Definition 1. *A copula C is a joint distribution function of standard uniform variables $U_1, U_2, ..., U_d$: $C(u_1, ..., u_d) = P(U_1 \leq u_1, ..., U_d \leq u_d)$ where each variable follows a standard uniform distribution $U_i \sim Uniform(0, 1)$ for $i = 1, 2, ..., d$.*

Sklar's Theorem states how the copula function relates to any joint distribution (see [14]). As a consequence of this, as shown in Equation (1) any d-dimensional joint density f and their marginal densities $f_1, f_2, ..., f_d$ are also related and can be represented as:

$$f(x_1, x_2, ..., x_d) = c(F_1(x_1), F_2(x_2), ..., F_d(x_d)) \cdot \prod_{i=1}^{d} f_i(x_i), \tag{1}$$

where c is the density of the copula C, X_i is a continous random variable, $F_i(x_i)$ is the marginal distribution function of X_i and $f_i(x_i)$ is the marginal density of

Table 1. Copula Density Functions.

Copula	Density Function	Parameter
Frank	$c(u_1, u_2, \theta) = \frac{-\theta(e^{-\theta}-1)e^{-\theta(u_1+u_2)}}{((e^{-\theta u_1}-1)(e^{-\theta u_2}-1)+(e^{-\theta}-1))^2}$	$\theta \in (-\infty, \infty) - \{0\}$
Gaussian	$c(u_1, u_2, \theta) = (1-\theta^2)^{-\frac{1}{2}} \exp\left(\frac{w_1^2+w_2^2-2\theta w_1 w_2}{2(1-\theta^2)} - \frac{w_1^2+w_2^2}{2}\right)$ where $w_1 = \Phi^{-1}(u_1)$ and $w_2 = \Phi^{-1}(u_2)$	$\theta \in (-1, 1)$

X_i. Equation (1) shows that the dependence structure is modeled by a copula function and that the marginal densities can have different distributions.

Bivariate Copula Functions In this paper we propose the use of two copula functions: Frank and Gaussian, for modeling two-dimensional copulas. Both of the copula proposed are symmetric, and they can model both positive and negative dependencies and their respective joint distributions have equally low and high values[1]. Table 1 shows the density functions of the two copula:

Where $u_1 = F_1(x_1)$ and $u_2 = F_2(x_2)$ are the transformed values of the features x_1 and x_2 through their marginal distribution functions. Φ^{-1} is the quantile function of a normal distribution. θ is the dependence parameter. The estimation of θ is done through the maximum likelihood method.

2.2 Probabilistic Classifier with Copula Functions

Starting from the the Naive Bayes classifier equation, we can add information of the association between features using copulas. The probabilistic classifier would be as shown in Equation (2):

$$P(A|x_1, ..., x_d) = \frac{c(F_1(x_1|A), ..., F_d(x_d|A))P(A) \times \prod_{i=1}^{d} f_i(x_i|A)}{f(x_1, ..., x_d)}, \qquad (2)$$

where A represent one of the classes, $x_1, x_2, ..., x_d$, are the values of the features, and c is the joint density copula function.

3 Case Study

The data available for this paper comes from a collaboration with the Escuela Nacional de Estudios Superiores Unidad León, UNAM. Prior to their participation in the study, all participants provided informed consent after being thoroughly briefed on the study's protocol, which is non-invasive. Additionally, this research adheres to the ethical principles outlined in the Helsinki Declaration. A total of 58 records of individuals performing mandibular opening/closing, lateral and protrusion/retrusion movements were used. Of the 58 records, 29 come from individuals diagnosed with temporomandibular

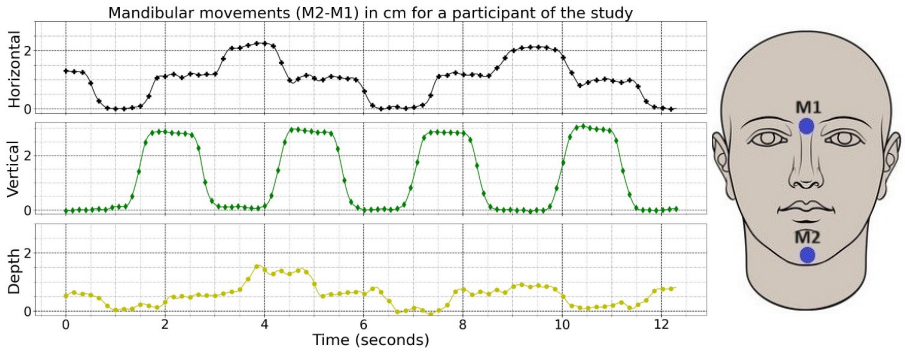

Fig. 1. Signals obtained from marker tracking.

arthralgia disorder and 29 records come from healthy individuals. For the clinical evaluation of the patient the DC-TMD[13] was used.

The records were obtained by tracking markers on the person's face. Two markers were placed in each person. Marker 1 was placed at the nasion craniometric point (between the eyes). Marker 2 was placed at the gnathion craniometric point (the chin). Marker 2 will record mandibular movement and marker 1 is a control marker for monitoring head movement. The equipment used for recording is an Intel Realsense Depth Camera D435. Tracking of the markers was carried out using the *OpenCV* library in Python.

The output of the tracking algorithms is a record of three signals per marker, corresponding to the three-dimensional position of the markers during mandibular movements. That is, each of the signals, shows us how the position of the marker changes with respect to the vertical, horizontal or depth axis over time. Figure 1 on the left shows segments of the signals obtained by tracking marker 2 for an individual. On the right it shows the location of marker 1 (M1) and marker 2 (M2) on the face as they were placed on the participants of the study.

4 Methodology

4.1 Preprocessing and Feature Extraction

The preprocessing consist of several stages: Initially a correction is applied to the signal to attenuate head movement artifacts from the mandibular movements. This is achieved by substracting the marker 1 signals from marker 2. Following this correction, the raw data from each recorded individual undergoes filtering using a 1D Gaussian filter and baseline correction. Subsequently the signals are segmented into three parts corresponding to the three mandibular movements recorded: Opening/closing (OC), Lateral (LAT) and Protrusion/Retrusion (PROT).

Table 2. Extracted features employed for TMD classification.

Biomechanical Features		
Name	Description	
OCY,OCX, OCZ	Max. deviation in y, x or z axis direction in opening movements	
LRX, LLX,LY,LZ	Max. deviation in x, y or z axis direction in lateral movements	
	Deviation in x axis measured to the left (LLX) and right (LRX)	
PY, PX, PZ	Maximum y, x or z axis deviation in Protrusion movements	
Speed, Acceleration and Jerk Features		
Name	Description	Formula
$\dot{s}_{max}$	Max speed	$\frac{\Delta s}{\Delta t}$
$\dot{s}_{avg}$	Mean speed	$\frac{\sum \dot{s}}{\lvert\dot{s}\rvert}$
$\sigma(\dot{s})$	Speed standard deviation	$\sqrt{\frac{\sum(\dot{s}-\dot{s}_{avg})^2}{\lvert\dot{s}\rvert}}$
$\ddot{s}_{max}$	Max acceleration	$\frac{\Delta \dot{s}}{\Delta t}$
$\ddot{s}_{avg}$	Mean Acceleration	$\frac{\sum \ddot{s}}{\lvert\ddot{s}\rvert}$
$\sigma(\ddot{s})$	Acceleration standard deviation	$\sqrt{\frac{\sum(\ddot{s}-\ddot{s}_{avg})^2}{\lvert\ddot{s}\rvert}}$
$\dddot{s}_{max}$	Max jerk	$\frac{\Delta \ddot{s}}{\Delta t}$
$\dddot{s}_{avg}$	Mean jerk	$\frac{\sum \dddot{s}}{\lvert\dddot{s}\rvert}$
$\sigma(\dddot{s})$	Jerk standard deviation	$\sqrt{\frac{\sum(\dddot{s}-\dddot{s}_{avg})^2}{\lvert\dddot{s}\rvert}}$
Frecuency Features		
Name	Description	Formula
P_{tot}	Total power	$\sum P_i$
P_{avg}	Mean power	$\frac{\sum P_i}{M}$
F_{ratio}	Frequency ratio	$\frac{\sum_{fs/2}^{fs} P_i}{\sum_{0}^{fs/2} P_i}$

When examining temporomandibular movements, various authors have proposed different feature extraction methods, including measurements such as maximum mouth opening [2,7], speed [3], Fourier transform [15], among others. In this study a collection of features previously proposed in other works for addressing TMD is used. Table 2 provides a summary of the ones extracted from the available data.

The speed, acceleration, jerk and frequency features were extracted for each of the signal segments corresponding to each of the mandibular movements. That is, the 12 proposed speed, acceleration, jerk and frequency features are extracted for each of the three signals (x, y and z axis) and for each of the movements (OC, LAT and PROT). That sums up to 108 features, to which we added the ten biomechanical features giving a total of 118 features.

4.2 Feature Selection

A subgroup of features was selected with the objective of dimensionality reduction. In this case the data was divided by the labels, having two groups:

Arthralgia group (AG) and the healthy group (HG). For each of the features extracted, a Wilcoxon test was applied to the groups. The Wilcoxon test compares the means of two groups, the p-value obtained by applying the test is a measure of how similar the means of the groups are. The lower the p-value, the larger is the difference in the means of the groups. From the original 118 features, the 20 features with the lowest p-values of the test were selected, Algorithm 1 shows the process in detail.

Algorithm 1 Feature Selection Pseudocode

```
Separate data in two subgroups AG and HG corresponding to each class.
for feature in datasets do
    Apply Wilcoxon test for the subgroups.
    Save the test p-value
end for
Sort in ascending order all p-values from the tests.
Select the 20 features with lowest p-values.
```

4.3 Copula Classifiers

Bivariate copula functions can model dependencies between two variables, but not all possible pairs of features show strong dependencies, therefore identifying which ones are the most relevant is a challenge. To this end, for the copula classifiers a tree graphical model based on Kruskal's minimum spanning tree algorithm(MST)[9] is proposed. Given a matrix $M_{d\times d}$, where d represents the total number of features, (in this case $d = 20$), contains the maximum log-likelihood for the combination of the i-th and j-th variables, according to the following equation (3):

$$M_{i,j} = \ell(\theta|u_i, u_j) = \sum \log c(\theta|u_i, u_j). \tag{3}$$

Using the MST algorithm requires transforming the maximum likelihood values into costs, which can easily be done by multiplying them by -1. When applied, the MST algorithm will render a tree-shaped graphic model showing the strongest dependencies to be modeled with a determined dataset. Algorithm 2 details the procedure for obtaining this graphical model.

Algorithm 2 is similar to the one proposed by [4] in the way the graphical model is obtained but with a few differences. One of them is that the criteria for determining the tree structure is log-likelihood instead of mutual information. Another is that the features in this work are continuous. Finally, the most important difference is that the dependence is modeled with copula functions. An example of the graphical model rendered by Algorithm 2 is shown in Figure 2. Then, Equation (4) defines a Bayes classifier based on bivariate copula functions and Figure 2:

Algorithm 2 Pseudocode for obtaining the tree graphical model

1: Separate training data by class and for each class dataset:
2: Estimate the distribution functions $F_1, F_2, ..., F_d$ of the features, for example, by using gaussian kernels
3: Transform the features $x_1, x_2, ..., x_d$ to $u_1, u_2, ..., u_d$ via $u_i = F_i(x_i)$
4: **for** i `in` $\{1, 2, .., d-1\}$ **do**
5: **for** j `in` $\{i+1, i+2, .., d\}$ **do**
6: Estimate $\theta_{i,j}$ parameter via maximum likelihood and save $M_{ij} = \ell(\theta|u_i, u_j)$
7: **end for**
8: **end for**
9: Multiply the matrix M by -1
10: Feed the matrix M to the MST algorithm

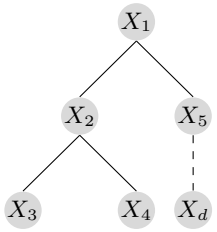

Fig. 2. Example of a dependence tree for a set of five features.

$$P(A|x_1, ..., x_d) = \frac{c(u_1, u_2)c(u_1, u_5)...c(u_{d-1}, u_d)P(A) \times \prod_{i=1}^{d} f_i(x_i|A)}{f(x_1, ..., x_d)}, \quad (4)$$

where $u_i = F_i(x_i)$ for $i = 1, 2, ..., d$. Then, going back to Figure 2, each of the edges connecting a pair of features represents a copula modeling the association between them. For the implementation of the copula classifiers the R package *MLCOPULA*[10] was used.

4.4 Model Training and Evaluation

A total of five models were trained for the classification task. Adding to the two copula models mentioned before (Frank and Gaussian), a Naive Bayes, KNN and Support Vector Machine were also trained. Each classifier is evaluated in randomly sampled partition cross validation scheme with 80% of the instances for training set and 20% for the validation set. Since the partitions are selected randomly, to ensure all instances are used in both training and validation, a high number of partition was proposed (50 partitions).

Four metrics were obtained for performance evaluation: accuracy, sensitivity, specificity and area under the ROC curve (AUC). The accuracy reflects the proportion of correctly classified instances, from both positive and negative classes; sensitivity shows the proportion of positive instances correctly classified,

Table 3. The 20 features selected via Algorithm 1 applied to the Features in Table 2.

OCX	OCY	OCZ	LRX
PZ	$OC\_Y\_\ddot{s}_{max}$	$OC\_Z\_\ddot{s}_{max}$	$OC\_Z\_\dddot{s}_{avg}$
$OC\_Y\_\dot{s}_{max}$	$OC\_Z\_\dot{s}_{max}$	$OC\_X\_\dot{s}_{avg}$	$OC\_Y\_P_{tot}$
$OC\_Y\_P_{ratio}$	$PROT\_Z\_\ddot{s}_{avg}$	$PROT\_Y\_\sigma(\dot{s})$	$PROT\_Z\_\dot{s}_{max}$
$PROT\_Y\_\dot{s}_{avg}$	$PROT\_Z\_\dot{s}_{avg}$	$PROT\_Z\_P_{avg}$	$LAT\_Y\_P_{tot}$

Table 4. Performance metrics by classifier. Best result in bold, standard deviation in parenthesis.

	Accuracy		Sensitivity		Specificity		AUC	
Classifier	Mean		Mean		Mean		Mean	
Naive Bayes	0.772	(0.12)	0.743	(0.20)	0.800	(0.17)	**0.833**	(0.13)
KNN	0.768	(0.12)	**0.800**	(0.16)	0.737	(0.17)	0.817	(0.12)
SVM	**0.782**	(0.10)	0.733	(0.15)	**0.830**	(0.16)	0.826	(0.10)
Frank	0.760	(0.11)	0.760	(0.21)	0.76	(0.17)	0.789	(0.14)
Gaussian	0.753	(0.11)	0.743	(0.20)	0.763	(0.18)	0.761	(0.14)

and specificity is the proportion of negative instances correctly classified. The AUC reflects the model's ability to distinguish between classes, the closer it is to 1, the better.

5 Results

The selected features using the method described in section 4 are shown in Table 3. The description of the features OCX, OCY, OCZ, LRX, PZ is shown in Table 2. To the rest of the features, the code OC, LAT and PROT represent the mandibular movement from which the feature was extracted: opening/closing, lateral and protrusion, respectively. The letter X, Y and Z represent the axis of the movement from which the feature was extracted: horizontal, vertical and depth, respectively.

Table 4 shows the mean and standard deviation of the measured metrics for each of the five classifiers tested. In bold is marked the best result for each of the metrics and in parenthesis the standard deviation. At first glance it is possible to see that although the performance of the copula classifiers is good overall, the SVM Classifier shows the best results in accuracy and specificity; the KNN had the best result for specificity and the Naive Bayes had the best AUC results.

Figure 3 shows the distribution of the accuracy and AUC for each of the classifiers. In terms of accuracy, the classifier with lower variance is the SVM, and is also the only one with atypical data points. The Gausssian copula, on the other hand, shows the highest variance of all the classifiers. Regarding the AUC, the SVM also shows the lowest variance and the Gaussian Copula the highest.

Figure 4 shows the distribution of the specificity and sensitivity for each of the classifiers. For the sensitivity metric, the boxplots show that the SVM

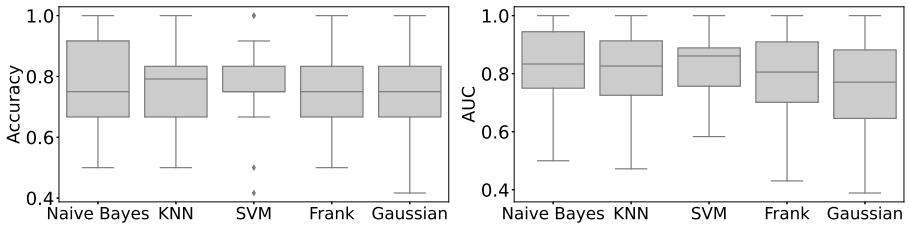

Fig. 3. Accuracy and AUC results by classifier.

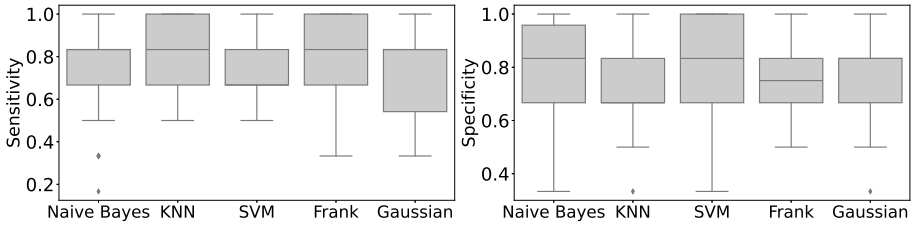

Fig. 4. Specificity and sensitivity results by classifier.

and Naive Bayes have the lowest variance, while the copula classifiers have the highest. The specificity boxplots, on the other hand, show the copula classifiers have the lowest variance.

To prove whether or not there is a difference in perfomance a Kruskal-Wallis test was applied to the results. Table 5 shows the p-values for the tests applied to the performance metrics. With a significance of $\alpha = 0.05$ the test shows that there are no differences in the means of the classifiers in the four metrics obtained.

6 Conclusions

This paper introduced copula classifiers with a tree graphical model for the classification of temporomandibular disorders. These classifiers were trained on movement data from healthy and TMD diagnosed individuals. Results show good mean performance in accuracy, sensitivity, specificity and AUC metrics for the copula classifiers. A comparison with other common classifiers used in similar tasks showed there is no statistical difference in performance for all the metrics used. Results also show that the copula classifiers have a higher variance in three out of the four metrics: accuracy, sensitivity and AUC.

Although the copula classifier performance is good, it was not particularly notable. This outcome may be related to the features selected for the training of the classifiers. Copula classifier perform best when there are strong relations

Table 5. Kruskal-Wallis test results for the performance metrics.

Metric	P-value	Metric	P-value
Accuracy	0.68130	AUC	0.07894
Sensitivity	0.3983	Specificity	0.05463

between the features in the data. Selecting the strongest dependencies among features is relevant in this proposed classifier because it could help differentiate between classes based on dependencies structures present in one class, but not in the other.

Since the selection method does not consider the strength of association among features, there is no guarantee that the selected features showed strong pairwise relationships. Another explanation for the performance is that the selected copulas do not provide the best model for the dependencies present in the data. Future work regarding copula classifiers for the prediction of temporomandibular disorders should consider a feature selection method accounting for feature dependencies. A set of copula functions modeling a wider range of dependencies should also be considered.

Acknowledgments. The first author Carlos López-Hernández acknowledges the financial support from the National Council of Humanities, Science and Technology of Mexico (CONAHCYT) given through the graduate scholarship program. The authors would like to thank Ulises Martín Arbaiza Martínez for applying the clinical evaluation of the participants in this study.

References

1. Balakrishna, N., Lai, C.D.: Continuous bivariate distributions. Springer Dordrecht Heidelberg London New York (2009)
2. Calil, B.C., Cunha, D.V.D., Vieira, M.F., Andrade, A.D.O., Furtado, D.A., Junior, D.P.B., Pereira, A.A.: Identification of arthropathy and myopathy of the temporomandibular syndrome by biomechanical facial features. BioMedical Engineering Online 19 (4 2020)
3. Castillo-Juárez, M.: Análisis cuantitativo de movimientos mandibulares para la clasificación de trastornos temporomandibulares por medio de un sistema de visión computacional. Tésis de Maestría, Universidad Autonoma de Aguascalientes (2022)
4. Chow, C., Liu, C.: Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory 14(3), 462–467 (1968)
5. Dowd, K.: Copulas in macroeconomics. Journal of International and global Economic Studies 1, 1–26 (7 2008)
6. Durham, J., Newton-John, T.R., Zakrzewska, J.M.: Temporomandibular disorders. BMJ (Online) 350 (3 2015)
7. Furtado, D.A., Pereira, A.A., Andrade, A.D.O., Peres, D., Junior, B., Silva, M.R.D.: A specialized motion capture system for real-time analysis of mandibular movements using infrared cameras (2013)

8. Genest, C., Favre, A.C.: Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering 12(4), 347–368 (2007)
9. Kruskal, J.B.: On the shortest spanning subtree of a graph and the traveling salesman problem. Proc. Amer. Math. Soc. 7, 48–50 (1956)
10. Salinas-Gutierres, R., Hernandez-Quintero, A., Montoya Calzada, A., Lopez Hernandez, C., Márquez Romero, J.: Mlcopula (2024), `https://cran.r-project.org/web/packages/MLCOPULA/`
11. Salinas-Gutiérrez, R., Hernández-Aguirre, A., Rivera-Meraz, M.J.J., Villa-Diharce, E.R.: Using Gaussian Copulas in Supervised Probabilistic Classification. Springer (2010)
12. Salinas-Gutiérrez, R., Hernández-Aguirre, A., Villa-Diharce, E.R.: Dependence trees with copula selection for continuous estimation of distribution algorithm. In: GECCO '11: Proceedings of the 13th annual conference on Genetic and evolutionary computation. pp. 585–592. ACM (2011)
13. Schiffman, E., Ohrbach, R., Truelove, E., Look, J., Anderson, G.: Diagnostic criteria for temporomandibular disorders (dc/tmd) for clinical and research applications: Recommendations of the international rdc/tmd consortium network * and orofacial pain special interest group † hhs public access. J Oral Facial Pain Headache 28, 6–27 (2014)
14. Sklar, A.: Fonctions de répartition á n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Universite de Paris 8, 229–231 (1959)
15. Taşkıran, U., Çunkaş, M.: A deep learning based decision support system for diagnosis of temporomandibular joint disorder. Applied Acoustics 182 (11 2021)
16. Troka, M., Wojnicz, W., Szepietowska, K., Podlasiński, M., Walerzak, S., Walerzak, K., Lubowiecka, I.: Towards classification of patients based on surface emg data of temporomandibular joint muscles using self-organising maps. Biomedical Signal Processing and Control 72, 103322 (2 2022)
17. Zhang, Y., Wang, X., Liu, D., Li, C., Liu, Q., Cai, Y., Yi, Y., Yang, Z.: Joint probability-based classifier based on vine copula method for land use classification of multispectral remote sensing data. Earth Science Informatics 13, 1079–1092 (2020)