Iranian Journal of Fuzzy Systems

Volume 22, Number 3, (2025), pp. 139-149

Original paper

University of Sistan and Baluchestan

A classification using mixture of concordance measures

A. Sheikhi [1] and R. Mesiar [2]

[1] *Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran*
[2] *Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, STU Bratislava, Slovakia*
[2] *UTIA AV ČR, Pod Vodárenskou věží 4, CZ-182 00, Prague 8 Czech Republic*

sheikhy.a@uk.ac.ir, radko.mesiar@stuba.sk

Abstract

In the realm of classification studies, existing literature indicates that, when the relationships among exploratory variables extend beyond linear functions, nonlinear classifiers tend to outperform their linear counterparts. This study employs concordance measures to attain optimal outcomes in a classification task. In this regard, we examine the connection copula among the exploratory variables, as well as the copula linking the exploratory attributes to the target attribute are taken into consideration. As a major novelty, our classification approach utilizes a convex combination of the pairwise Spearman's rank correlation coefficient ρ and the pairwise Kendall's association τ. Through a simulation analysis, we assess the performance of our algorithm, which demonstrates its superiority over alternatives, including copula-based classification methods as well as machine learning classification models. We also, provide an application of our method to the classification of COVID-19 dataset for more illustration.

Keywords: Concordance measure, copula, classification, correlation, association measure.

1 Introduction

As computer and software technologies continue to reach an impressive level, data scientists are striving to develop more efficient models that exhibit enhanced performance, particularly in the context of big data. The increasing complexity of these models can be handled through the advancements in computational technology [1]. This effort in classification models has been carried out in the academic literature. Chen et al.(2022) [22] have improved the accuracy of the classification process using a recursive feature elimination. Douider et al. (2024) [15] have designed a procedure to improve the performance of a binary classification. For more recent works about the advances of classification, one may refer to Che et al. (2023) [2] and Sun et al. (2024) [32] and references therein. Hou et al. (2024) [31] have introduced a model-based method for a nonlinear classification.

In this study, we aim to leverage copula functions to enhance classification accuracy, particularly in scenarios where attributes exhibit nonlinear relationships. Voisin et al. (2013) [29] have investigated a supervised classification with a hierarchical copula-based approach. Mesiar and Sheikhi (2021) [18] have introduced a nonlinear random forest classification algorithm by the aid of copula functions. Dong et al. (2017) [6] have performed an image classification using copula-based joint statistical models. For additional insights into the use of copula functions in classification algorithms, one may refer to [7], [8] and [11].

In the context of classification task using association measures, Han and Liu (2012) [9] have used a matrix containing pairwise Spearman's rank correlation coefficient ρ coefficients as an alternative of the Pearson correlation matrix in a semi-parametric principal component analysis, see also Han and Liu (2014) [10]. For more application of the correlation and association measures in the wide range of data science, we direct readers to the works of Toharudin et al. (2019) [21]. The main aim of this paper is leveraging the classification accuracy using combinations of concordance measures. The utilization of copulas and dependence measures within the field of data science has been thoroughly examined in

Corresponding Author: A. Sheikhi
Received: February 2025; Revised: June 2025; Accepted: June 2025.
https://doi.org/10.22111/ijfs.2025.51019.9017

the literature. See for instances, Snehalika et al. (2020) [13] for investigating copula-based feature selection methods; and Lascio et al. (2018) [4] for copula-based clustering approaches.

Our results showed the contribution of our proposed method in contrary to other classification alternatives which are applying copula functions. We also showed that our approach is superior to machine learning algorithms including support vector machine as well as random forest models.

The subsequent sections of the paper are organized as follows. We provide preliminary materials in the next section. Our method is introduced in section 3. Some numerical analysis including simulation study alongside with a real data application are given in section 4. Finally, section 5 contains some concluding remarks.

2 Copula and concordance measures

Assume that the random vector $\boldsymbol{X} = (X_1, X_2, ..., X_d)$ follows the joint multivariate distribution function $F_{X_1,X_2,...X_d} : \mathbb{R}^d \to [0,1]$ and $F_{X_i} : \mathbb{R} \to [0,1]$, $i = 1, 2, ..., d$, are the related marginal distribution functions of X_i, $i = 1, 2, ..., d$. Based on the Sklar's theorem [27] there exists a grounded, uniformly marginal and $d-$increasing $d-$variate function $C : [0,1]^d \to [0,1]$ such that relates the marginals to the joint distribution as

$$F_{X_1,X_2,...X_d}(x_1, x_2, ..., x_d) = C(F_{X_1}(x_1), F_{X_2}(x_2), ..., F_{X_d}(x_d)), \tag{1}$$

which is called copula function and is unique if all X_i, $i = 1, 2, ..., d$ are absolutely continuous random variables. In this case, from (1), the density h of $(X_1, .., X_d)$ can be represented as follows

$$h(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d)) \prod_{k=1}^{d} f_k(x_k), \tag{2}$$

where c is the density of the copula C and f_k, $k = 1, .., d$, are the respective densities of the random variables X_1, $.., X_d$. See, e.g., Nelsen (2006) [19], for more details. We assume $d = 2$, so (X, Y) is a continuous bivariate random vector and C is the related copula. In order to start our work, we denote by $\mathcal{C}$ the set of all 2-variate copulas. Then, for two related random variables X and Y related by a copula $C \in \mathcal{C}$, inspired by Scarsini (1984) [26], see also Mesiar et al. (2022) [16], we first adopt the following operators:
$v : \mathcal{C} \to \mathcal{C}, v(C)(x, y) = C(y, x)$;
$\varphi_1 : \mathcal{C} \to \mathcal{C}, \varphi_1(C)(x, y) = y - C(1 - x, y)$;
$\varphi_2 : \mathcal{C} \to \mathcal{C}, \varphi_2(C)(x, y) = x - C(x, 1 - y)$;
$\varphi : \mathcal{C} \to \mathcal{C}, \varphi(C)(x, y) = x + y - 1 + C(1 - x, 1 - y)$,
which are respectively called permutation, x-reflection (first flipping), y-reflection (second flipping) and reversion (survival operation). Observe that the independence copula Π is invariant under any of the above introduced involutive transformations. The same is true for the Fréchet copulas $aM + aW + (1-2a)\Pi, a \in [0, 0.5]$, in particular, for $a = \frac{M+W}{2}$ [26]. The following definition investigates the properties of a so called concordance measure.

Definition 2.1. *A measure of concordance is a function $\kappa : \mathcal{C} \to [-1, 1]$ that satisfies the following properties:*
(1) $\kappa(C) = \kappa(\nu(C))$;
(2) $\kappa(C_1) \leq \kappa(C_2)$ *whenever* $C_1 \prec C_2$;
(3) $\kappa(\varphi_1(C)) = -\kappa(C)$;
(4) $\kappa(\Pi) = 0$ *and* $\kappa(M) = 1$;
(5) *if* $C_n \underset{n\to\infty}{\longrightarrow} C$, *then* $\lim_{n\to\infty} \kappa(C_n) = \kappa(C)$,
where $\prec$ denotes the concordance order.

For more details concerning copulas, their families and association measures we recommend Nelsen (2006) [19] and Dolati and Úbeda Flores (2006) [5]. In this work we use the convex combination of concordance measures to carry out a classification task. We use a convex combination of Spearman's rank correlation coefficient ρ and the pairwise Kendall's association τ in a classification algorithm. The utilization of copulas and dependence measures within the field of data science has been thoroughly examined in the literature. For some works in this topic, one may refer to Salinas-Gutiérrez et al. (2010) [24] who have made significant contributions by employing Gaussian copula within a supervised classification framework and Ozdemir et al. (2017) [20] who explored various classification algorithms that incorporate copulas. Also in this context, Chen (2016) [3] has introduced a copula-based classification for continuous and discrete data. In this work we use the convex combination of concordance measures to carry out a classification task. Consider the distribution functions of X_1, Y_1 and X_2, Y_2 are continuous. Based on Kruskal [12], the concordance

function is defined for a pair of random vectors X_1, Y_1 and X_2, Y_2. If X_1, Y_1 are associated to copula C_1 and C_2 is the connecting copula of X_2 and Y_2, then their concordance function is defined as

$$Q_{X,Y} = Q(C_1, C_2) = 4\int_0^1 \int_0^1 C_2(u,v)dC_1(u,v) - 1. \tag{3}$$

It is known that the concordance measure (3) is related to several correlations as well as association measures, among them Kendall's τ and Spearman's ρ. More specific, denoting $C = C_{X,Y}$, (3) yields Kendall's measure of association $\tau_{X,Y}$ which is defined as

$$\tau_{X,Y} = Q(C, C) = 4\int_0^1 \int_0^1 C(u,v)dC(u,v) - 1.$$

Similarly, considering $C_2 = \Pi$, Spearman's measure of association $\rho_{X,Y}$ is given by

$$\rho_{X,Y} = 3Q(C, \Pi) = 12\int_0^1 \int_0^1 C(u,v)dudv - 3.$$

For more details concerning copulas, their families and association measures we recommend Nelsen (2006) [19] and Dolati and Úbeda Flores (2006) [5]. In the sequel, for the sake of the notational simplicity, we adopt the notations τ, ρ for $\tau_{X,Y}$ and $\rho_{X,Y}$ respectively and we use a convex combination of them in a classification algorithm.

3 Classification based on the concordance measure

The first aim of this section is investigating the estimator of the weight in convex combination of two estimations. To do this, the following lemma will play a significant role.

Lemma 3.1. *Let $\boldsymbol{X} \sim C(\boldsymbol{u}, \boldsymbol{\theta})$, where $\boldsymbol{u} = (u_1, ..., u_d) \in [0,1]^d$ is the vector of copula observations and $\boldsymbol{\theta}$ is the vector of its parameters. Assume that $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are two estimates of the copula parameter $\boldsymbol{\theta}$ and we let $\boldsymbol{\theta}(\alpha) := \alpha\hat{\boldsymbol{\theta}}_1 + (1-\alpha)\hat{\boldsymbol{\theta}}_2$, $\alpha \in [0,1]$, then the MLE of α is obtained from the equation*

$$\sum_{i=1}^{n} \frac{1}{c\left(\boldsymbol{u}_i; \boldsymbol{\theta}(\alpha)\right)} \frac{\partial}{\partial \boldsymbol{\theta}(\alpha)} c\left(\boldsymbol{u}_i; \boldsymbol{\theta}(\alpha)\right) \frac{\partial}{\partial \alpha}\boldsymbol{\theta}(\alpha) = 0, \tag{4}$$

where $\boldsymbol{u}_i = (u_{i1}, ..., u_{id}), i = 1, ..., n$ is the i-th realization of n-tuple of copula observations.

Proof. Via the assumption of standard uniform of marginals $u_i,\ i = 1, ..., d$, we readily construct the likelihood function $\mathcal{L}(\alpha) = \prod_{i=1}^{d} c(\boldsymbol{u}_i, \boldsymbol{\theta}(\alpha))$. So, the proof is straightforward using a logarithmic transform of $\mathcal{L}(.)$ and a chain rule derivation. □

The solution of the equation (4) is dependent on the type of copula. The following two corollaries consider 2-dimensional Farlie-Gumbel-Morgenstern (FGM) and Gaussian copulas; denoting $\hat{\boldsymbol{\theta}}_i = \hat{\theta}_i,\ i = 1, 2$. The density of a FGM copula with parameter θ is denoted by

$$c(u, v; \theta) = 1 + \theta(1-2u)(1-2v), \quad u, v \in [0,1], \tag{5}$$

and the density of a 2-dimensional Gaussian copula with correlation parameter θ is given by:

$$c(u, v; \theta) = \frac{1}{\sqrt{1-\theta^2}} \exp\left(-\frac{\theta^2(s^2+t^2) - 2\theta st}{2(1-\theta^2)}\right), \tag{6}$$

where for $u, v \in [0,1],\ s = \Phi^{-1}(u)$ and $t = \Phi^{-1}(v)$ are the standard normal quantiles.

Corollary 3.2. *Let $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of parameter θ in a 2-dimensional FGM copula and $\hat{\theta}(\alpha) = \alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2,\ 0 \le \alpha \le 1$, then an approximation of optimal weight α is*

$$\hat{\alpha} \approx \frac{\sum_{i=1}^{n}(1-2u_i)(1-2v_i) - \sum_{i=1}^{n}(1-2u_i)^2(1-2v_i)^2\hat{\theta}_2}{\sum_{i=1}^{n}(1-2u_i)^2(1-2v_i)^2\hat{\theta}_1 - \sum_{i=1}^{n}(1-2u_i)^2(1-2v_i)^2\hat{\theta}_2}.$$

Proof. From (5), the log-likelihood for a sample of size n is

$$\ell(\theta) = \sum_{i=1}^{n} \ln\left(1 + \theta(1-2u_i)(1-2v_i)\right).$$

So, by expanding the derivation of the log-likelihood function using the Taylor expansion we obtain

$$\begin{aligned}\frac{\partial \ell(\alpha)}{\partial \alpha} &= \sum_{i=1}^{n} \frac{(1-2u_i)(1-2v_i)}{1+\hat{\theta}_\alpha(1-2u_i)(1-2v_i)} \frac{\partial \hat{\theta}_\alpha}{\partial \alpha} \\ &= \sum_{i=1}^{n}(1-2u_i)(1-2v_i)\frac{\partial \hat{\theta}_\alpha}{\partial \alpha} - \sum_{i=1}^{n}(1-2u_i)^2(1-2v_i)^2\frac{\partial \hat{\theta}_\alpha}{\partial \alpha}\hat{\theta}_\alpha + O(\hat{\theta}_\alpha^2).\end{aligned}$$

For solving $\frac{\partial \ell(\alpha)}{\partial \alpha} = 0$, we note that for different $\hat{\theta}_1$ and $\hat{\theta}_2$, $\frac{\partial \hat{\theta}_\alpha}{\partial \alpha} = (\hat{\theta}_1 - \hat{\theta}_2) \neq 0$, Hence, an approximation of $\hat{\alpha}$ can be obtained as

$$\hat{\alpha} \approx \frac{\sum_{i=1}^{n}(1-2u_i)(1-2v_i) - \sum_{i=1}^{n}(1-2u_i)^2(1-2v_i)^2\hat{\theta}_2}{\sum_{i=1}^{n}(1-2u_i)^2(1-2v_i)^2\hat{\theta}_1 - \sum_{i=1}^{n}(1-2u_i)^2(1-2v_i)^2\hat{\theta}_2}.$$

□

The following corollary need to be solved numerically in the case of gaussian copula.

Corollary 3.3. *Let $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of parameter θ in a 2-dimensional Gaussian copula and $\hat{\theta}(\alpha) = \alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2,\ 0 \leq \alpha \leq 1$, then the optimal weight α is obtained via solving*

$$\frac{n\hat{\theta}_\alpha}{1-\hat{\theta}_\alpha^2} + \sum \frac{2[\hat{\theta}_\alpha(s_i^2+t_i^2) - s_i t_i] - \hat{\theta}_\alpha^2 s_i t_i}{(1-\hat{\theta}_\alpha^2)^2} = 0.$$

Proof. Similar to the prove of corollary (3.2), the log-likelihood function for a sample of size n is

$$\ell(\alpha) = \sum_{i=1}^{n} \ln c(u_i, v_i; \theta_\alpha) \quad = \sum_{i=1}^{n}\left[-\frac{1}{2}\ln(1-\hat{\theta}_\alpha^2) - \frac{\hat{\theta}_\alpha^2(s_i^2+t_i^2) - 2\hat{\theta}_\alpha s_i t_i}{2(1-\hat{\theta}_\alpha^2)}\right],$$

again knowing that $\hat{\theta}_1 - \hat{\theta}_2$, is non zero for different estimators, and after some algebra calculations, from $\frac{\partial \ell(\alpha)}{\partial \alpha} = 0$ we obtain the equation

$$\frac{n\hat{\theta}_\alpha}{1-\hat{\theta}_\alpha^2} + \sum \frac{2[\hat{\theta}_\alpha(s_i^2+t_i^2) - s_i t_i] - \hat{\theta}_\alpha^2 s_i t_i}{(1-\hat{\theta}_\alpha^2)^2} = 0.$$

□

Let $\boldsymbol{m} = \{m_1, \ldots, m_k\}$ be a finite set of k levels of classes. As a matter of facts, these k levels correspond to a target attribute, let us say y. Also, assume that the exploratory feature space be $\boldsymbol{x} = (x_1, ..., x_d) \in \mathbb{R}^d$. The problem of classification is to assign each $\boldsymbol{x}$ a label $m_1, \ldots, m_k$. Let $g(.)$ be a discriminant function, and for each of class levels, consider $g_i(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R},\ i = 1, ...,\ k$. Then g assigns an observation $\boldsymbol{x}$ a label from m_i if,

$$g_i(\boldsymbol{x}) > g_j(\boldsymbol{x}) \quad \text{for all } j \neq i.$$

Evidently, from (2) and in line with Bayes's theorem we define such discriminant functions as follows

$$\begin{aligned} g_i(x_1, \ldots, x_d) = P(m_i \mid x_1, \ldots, x_d) &= \frac{h(x_1, \ldots, x_d \mid m_i) P(m_i)}{h(x_1, \ldots, x_d)} \propto h(x_1, \ldots, x_d \mid m_i) P(m_i) \\ &= c(F_1(x_1), \ldots, F_d(x_d), \theta \mid m_i) \prod_{j=1}^{d} f_j(x_j, \theta_j \mid m_i) P(m_i) \\ &= c_i(F_1(x_1), \ldots, F_d(x_d), \theta_i) \prod_{j=1}^{d} f_j(x_j, \theta_{i,j}) P(m_i), \quad i = 1, 2, \ldots, k, \end{aligned}$$

where $h(.\mid m_i):\mathbb{R}^d\to\mathbb{R}^+$is a likelihood function and $P(m_i), i=1,\ldots,k$, are prior probabilities of the classes from $\boldsymbol{m}$. Also, $\theta_i, i=1,2,\ldots,k$, is the connection parameter of copula in the i-th class and $\theta_{i,j}, i=1,2,\ldots,k$, and $j=1,\ldots,d$, are the parameters of the j-th marginal distributions in the i-th class in which all these connection and marginal parameters are in the parameter space Θ, namely, $\Theta=\{\theta_i, i=1,2,\ldots,k \text{ and } \theta_{i,j}, i=1,2,\ldots,k, j=1,\ldots,d\}$. Moreover, $\propto$ indicates the proportionality of the left side to the right, because $h(x_1,\ldots,x_d)$ is the same for all classes.

Finally, in order to avoid the situation of $g_i(\boldsymbol{x})=g_j(\boldsymbol{x})$, for any $j\neq i$, we ignore tie observations if exist. For each class $m_i, i=1,\ldots,k$, let N_i represents training samples from the i^{th} class and define $\mathcal{L}_i(\boldsymbol{x},\theta)$ to be the likelihood function for these samples. Then we have

$$\mathcal{L}_i(\boldsymbol{x},\theta)=\prod_{n=1}^{N_i}\left[c_i\left(F_1\left(x_1^{(n)}\right),\ldots,F_d\left(x_d^{(n)}\right),\theta_i\right)\prod_{j=1}^{d}f_j\left(x_j^{(n)},\theta_{i,j}\right)\right], i=1,...,k, \tag{7}$$

where c_i is the copula density for the i-th class. In order to present our classification approach, without loss of generality we assume $d=2$. We first find a convex coefficient $\hat{\alpha}_i$ for each class $m_i, i=1,\ldots,k$, with respect to (7) with the objective function

$$\hat{\alpha}_i=\underset{\alpha_i\in[0,1]}{\arg\max}\left\{\mathcal{L}_i\left(\boldsymbol{x},\alpha_i\hat{\rho}_i+(1-\alpha_i)\hat{\tau}_i\right)\right\},\quad i=1,2,\ldots,k, \tag{8}$$

where $\hat{\rho}_i$ and $\hat{\tau}_i$ are the estimated values of Spearman's rank correlation coefficient ρ and the pairwise Kendall's association τ respectively in the i-th class. Once the estimated value of $\hat{\alpha}_i$ was obtained via (8), we define the estimation of our concordance measure in the i-th class as a function of $\hat{\alpha}_i$ as

$$\hat{\theta}_i(\hat{\alpha}_i)=\hat{\alpha}_i\hat{\theta}_{1i}+(1-\hat{\alpha}_i)\,\hat{\theta}_{2i},\quad i=1,2,\ldots,k, \tag{9}$$

where $\hat{\theta}_{1i}=\hat{\rho}_i,\ \ \hat{\theta}_{2i}=\hat{\tau}_i$. The following proposition claims the existence of such $\hat{\alpha}_i$.

Proposition 3.4. *There exist an $\hat{\alpha}_i\in[0,1], i=1,2,\ldots,k$, in which (9) is a concordance measure.*

Proof. We need to prove that (9) satisfies properties of the Definition 1. Properties (1),.., (4) are obvious via the permutation, x_reflection, y_reflection and reversion properties of ρ_i and τ_i. Also, from the convexity of (6) with respect to $\hat{\alpha}_i$, and knowing that if $C_n\underset{n\to\infty}{\longrightarrow}C$ then $\lim_{n\to\infty}\tau_i(C_n)=\tau_i(C)$ and $\lim_{n\to\infty}\rho_i(C_n)=\rho_i(C)$, we deduce that $\lim_{n\to\infty}\theta_i(C_n)=\theta_i(C)$. □

Although in Proposition 3.4 we proved the existence of such α and we obtained the optimal values of such α in Corollaries 3.2 and 3.3, however, as a matter of facts, our approach can be seen as an exploratory approach which help us to find the optimal values of the likelihood functions, which yields a better classification. In the sequel, we substantiate this method.

Wong et al. (2020) [30] have introduced a high performance hybrid classification algorithms which is based on instance filtering. In order to implement (8), we modify the algorithm of Tsung et al. (2020) [30] as follows. (See also, Algorithm 3). We divide the dataset into two distinct subsets: the training set and the testing set. The hybrid classification process, which is predicated on instance filtering, consists of three stages (refer to Figure 1). The initial stage involves the implementation of a fundamental classification algorithm, namely L_1, on the training set. Potential classification algorithms for L_1 include random forest, support vector machine, naive Bayes, and K nearest neighbors, among others.

As illustrated in the upper-right dashed rectangle of Figure 1, the training instances that are accurately classified through a k-fold cross-validation of the algorithm L_1 are allocated to a "primary" set, while those that are misclassified are stored in a "secondary set. Let us consider the data set D partitioned into two non-overlapping subsets: the training set D_R and the testing set D_T. Upon implementing algorithm L_1 on the training set D_R, which is subsequently assessed through k-fold cross-validation, we obtain a primary subset D_{RP} and a secondary subset D_{RS}. It is important to note that these subsets satisfy the conditions $D_{RP}\cap D_{RS}=\emptyset$ and $D_{RP}\cup D_{RS}=D_R$. Then, on each of these three sets D_R, D_{RP} and D_{RS} we perform our algorithm based on convex combination of concordance measures Spearman's ρ and Kendall's τ. This novel approach expands the accuracy of the considered models.

We call the outcomes associated with these three sets as the "full set", the "primary set", and the "secondary set", represented by $M_{D_{R,L2}}, M_{D_{RP,L2}}$ and $M_{D_{RP,L2}}$, respectively. This is illustrated in the lower-right broken rectangle of Figure 1. The left broken rectangle in the same figure depicts the third step, wherein each of these three models will be utilized to forecast the class value for each instance within the testing set. The models induced from the full training set, the primary set, and the secondary set by classification algorithm L_2 are called the full set, the primary set, and the secondary set, respectively in the second step, see the lower-right broken rectangle in Figure 1. The potential

classification algorithms for L_2 are the same as L_1. The left broken rectangle in the same figure shows the third step, in which one of the three sets will be chosen to predict the class value of each instance from the testing set.

The degree of similarity between a testing instance $\boldsymbol{x} = \langle x_1, x_2, \ldots, x_n \rangle \in D_T$ and the dataset D_{Rm} is quantified by the probability $p\left(D_{Rm} \mid \boldsymbol{x}\right) = p\left(\boldsymbol{x} \mid D_{Rm}\right) p\left(D_{Rm}\right) / p(\boldsymbol{x})$, where m can be either P or S. Here, n denotes the total number of attributes present in the dataset D. It is assumed that the attributes are independent within any specified set D_{Rm}. Under this assumption, the expression for $p\left(D_{Rm} \mid \boldsymbol{x}\right)$ can be further simplified as

$$p\left(D_{Rm} \mid \boldsymbol{x}\right) \propto p(D_{Rm}) \mathcal{L}\left(\boldsymbol{x}, \theta_{Rm}\right).$$

Let $|D|$ denote the number of instances in set D. Then $p\left(D_{Rm}\right)$ is estimated as $|D_{Rm}| \,/\, |D_R|$. By estimating $\hat{\theta}_{RP}$ and $\hat{\theta}_{Rs}$ in sets D_{RP} and D_{RS} respectively, we assign a testing instance $\boldsymbol{x}$ into D_{RP} if

$$log_2 p\left(D_{RP} \mid \boldsymbol{x}\right) - log_2 p\left(D_{RS} \mid \boldsymbol{x}\right) > \delta,$$

else, $\boldsymbol{x}$ will belong to D_{RS} where δ is a threshold parameter.

Following Wong et al. (2020)[30], in order to handle the case when $\mathcal{L}(\boldsymbol{x}, \hat{\theta}_{Rm})$ and $\mathcal{L}(\boldsymbol{x}, \hat{\theta}_{RS})$ are close, model $M_{DR,L2}$ can be a good choice for specifying the class value of $\boldsymbol{x}$, so, we classify $\boldsymbol{x}$ with $M_{DR,L2}$ if

$$log_2 p(D_{RP} \mid \boldsymbol{x}) - log_2 p(D_{RS} \mid \boldsymbol{x}) < \delta.$$

An algorithm of this approach is presented in Algorithm 3.

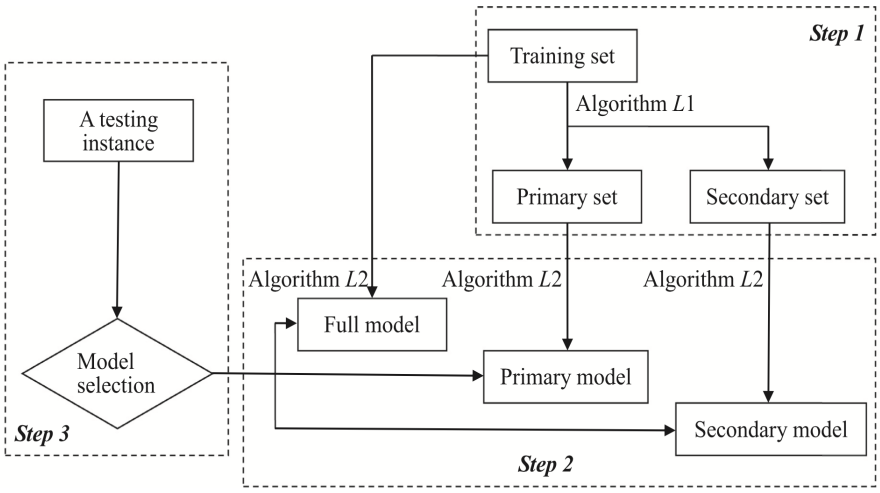


Figure 1: Classification based on filtering algorithm [30].

The confusion matrix serves as a crucial metric for evaluating the performance of classification algorithms. It commonly referred to as a contingency table which takes the form of a 2×2 table in the context of binary classification, as illustrated in Figure 2. The metrics of Sensitivity, Specificity and Accuracy were our criteria to asses the performance of our algorithm. Sensitivity quantifies the effectiveness of a classification task in recognizing true positives, whereas Specificity assesses its ability to identify true negatives. Accuracy, on the other hand, evaluates the overall correctness of the model's predictions. More specifically,

$$Sensitivity = \frac{TP}{TP + FN}, \quad Specificity = \frac{TN}{TN + FP}, \quad Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP, TN, FP, and FN are defined as Figure 2.

4 Numerical results

In this section, demonstrations of our method is provided through both a simulated data set and an analysis of real data.

Algorithm 1: Classification using a convex combination of concordance measures

Data: Data set $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{y})$, where $\boldsymbol{X}$ is the matrix of exploratory variable and $\boldsymbol{y}$ includes observations of class variable, threshold value δ.

Result: Selected convex combination of concordance measures, Classification based on a modification of the filtering algorithm.

Initialization : Perform a k-fold cross validation technique to split the set D into folds $F_1, F_2, \ldots, F_k$.
for *each of testing fold* F_j **do**
  Assign $D_T = F_j$ and $D_R = D \backslash F_j$;
  Execute algorithm L_1 on the set D_R to generate D_{RP} and D_{RS};
  Execute algorithm L_2 on the sets D_R, D_{RP} and D_{RS} to obtain models;
  $M_{D_{RP},L2}$, $M_{D_{RS},L2}$, and $M_{D_R,L2}$;
  For each instance $\boldsymbol{x} \in D_T$;
  Estimate $\hat{\rho}$ and $\hat{\tau}$ for the observations of $(X_i,\ X_j)$, $i, j = 1, 2, ..., p,\ i \neq j$;
  Select the optimum value of α using 8;
  Based of the value of α, calculate $p(D_{RP} \mid \boldsymbol{x})$ and $p(D_{RS} \mid \boldsymbol{x})$;
  If $\log_2 p(D_{RP} \mid \boldsymbol{x}) - \log_2 p(D_{RS} \mid \boldsymbol{x}) > \delta$, classify $\boldsymbol{x}$ using the model $M_{D_{RP},L2}$;
  Else-if, $\log_2 p(D_{RS} \mid \boldsymbol{x}) - \log_2 p(D_{RP} \mid \boldsymbol{x}) > \delta$, classify $\boldsymbol{x}$ using model $M_{D_{RS},L2}$;
  Else, use the model $M_{D_R,L2}$ to classify $\boldsymbol{x}$.
end

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positives (TP)	False Negatives (FN)
	Negative	False Positives (FP)	True Negatives (TN)

Figure 2: Confusion matrix

4.1 Simulation study

In order to perform a simulation study we generate data from Gaussian copula. We generate $n = 1000$ random samples (X_1, Y_1) from Gaussian copula with $\rho = 0.85$. Following the algorithm 3 we first estimate $\hat{\rho}$ and $\hat{\tau}$ for the observations of random vector $(X_1,\ X_2)$. We compute likelihood function of (7) for value of $\hat{\rho}$, value of $\hat{\tau}$ and value of $\hat{\theta}$ where values of α varies form 0 to 1 with steps of 0.05. A scatter plot as well as the behavior of their joint distribution for the different values of the α is depicted in the Figure 3 (panels (a) and (b)). To investigate the effect of concordance measures in detail, we produce inverse, square and exponential transforms of X_1, namely X_2, X_3 and X_4, respectively. In a similar fashion, we obtain the maximum log likelihood alongside with scatter plots for the pairs $(X_1,\ X_2)$, $(X_1,\ X_3)$ and $(X_1,\ X_4)$. See panels (c) and (d) of the same figure for the corresponding plots for the random vector $(X_1,\ X_2)$, panels (e) and (f) for the random vector $(X_1,\ X_3)$ and panels (g) and (h) for $(X_1,\ X_4)$. Additionally, as an alternative scenario, for the last row of this figure, we assume that the parameter of the Gaussian couple is equal to 0 which means the dependence of the generated data. Indeed, we generate X_5 as an independent random variable of X_1, refer to the fifth row of panels of Figure 3. As can be seen from this figure, for the inverse transformation the behavior of the log likelihood function is a squared form and its maximum value occurred when α equals 0.5. Also, for the squared function the values of α happens at 0.88, and in all other cases the best value of α will be 1. This information will lead us to make a better classification.

Regarding to obtain a class valued variable, we apply a convex function of two generated columns of the data and obtain an outcome variable by recoding the negative values of the outcome variable "0" and other values "1". Utilizing the estimated concordance measures $\hat{\rho}$, $\hat{\tau}$, and $\hat{\theta}$ as the dependency parameters, we have in mind Algorithm 3 to classify the target values. The means of outcomes of the classification criteria are summarized in Table 1, which shows the superiority of the convex combination of concordance measure contrary to $\hat{\rho}$, $\hat{\tau}$ as well as the independence

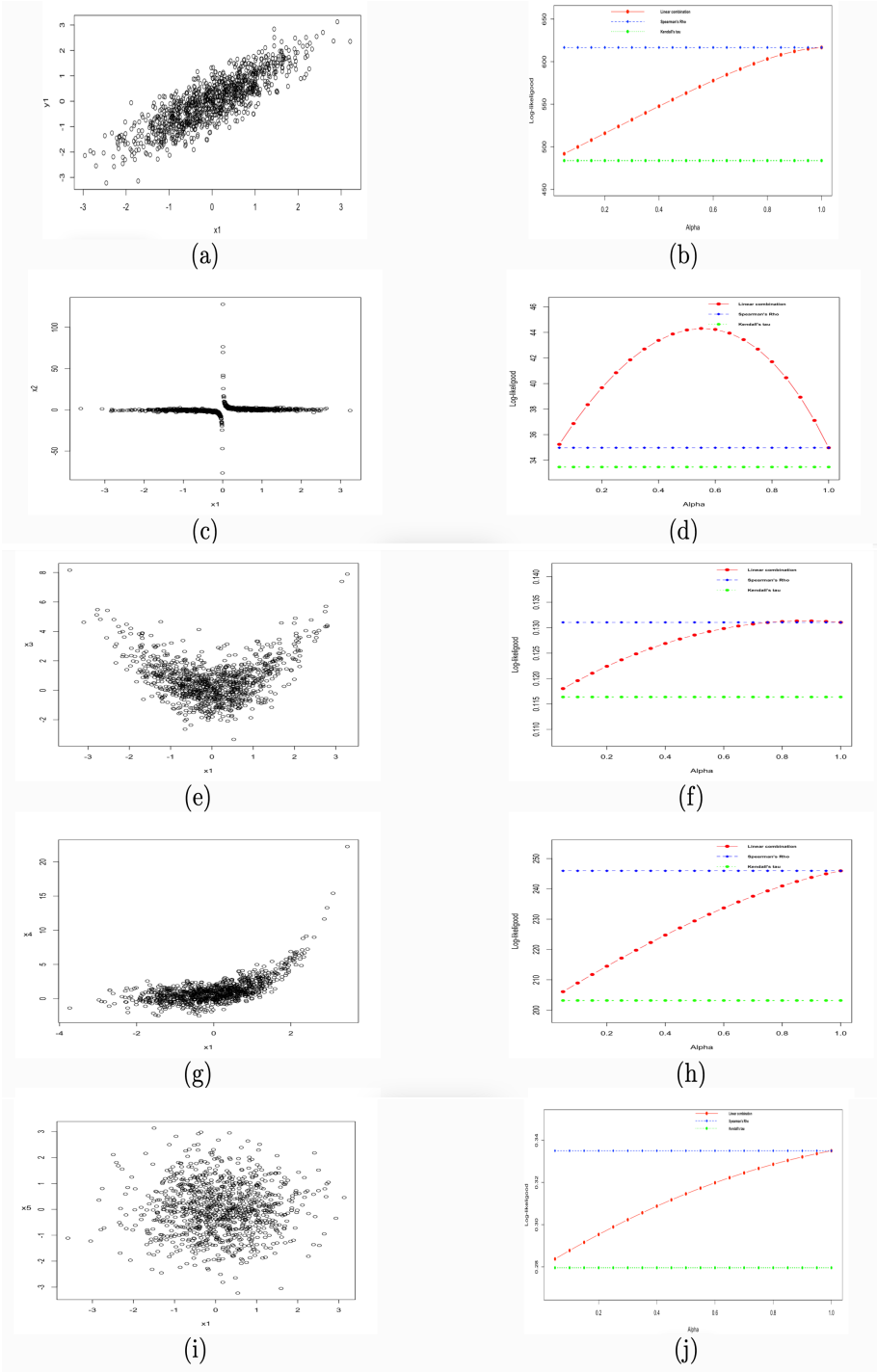


Figure 3: Scatter plot of the simulated data and the behavior of the log-likelihood.
First row: Gaussian copula. *Second row:* Inverse function. *Third row:* Square function. *Fourth row:* Exponential function. *Fifth row:* Independent samples.

assumption. In addition, to assess the performance of our algorithm, we compare the result of our classification with other alternatives. Regarding to compare our contributions with copula-based classification techniques, we use the approach of MLCOPULA of Salinas-Gutiérrez et al. (2025) [23], see also, [25]. We also apply a traditional Gaussian copula classification as another alternative to classify the outcome variable [24]. Moreover, we consider machine learning models as other alternatives. Choosing the support vector machine (SVM) and the random forest (RF) techniques, we

classify the outcome variable. The results are summarized in Table 1. As this table reveals, the concordance approach is superior to all other, while we notice that there is a close competition between our approach and the MLCOPULA algorithm in the specificity criterion.

Table 1: Accuracy of classification in simulated data

Approach	Parameter of dependency	Sensitivity	Specificity	Accuracy
Concordance ($\hat{\theta} = \hat{\rho}$)	Spearman's ρ	0.788	0.783	0.791
Concordance ($\hat{\theta} = \hat{\tau}$)	Kendall's τ	0.811	0.820	0.832
Concordance ($\hat{\theta} = \hat{\theta}_{cc}$)	convex concordance	**0.903**	**0.837**	**0.889**
Concordance ($\hat{\theta} = 0$)	independent attributes	0.706	0.828	0.768
MLCOPULA	Pearson's ρ	0.901	**0.837**	0.869
Gaussian copula	Pearson's ρ	0.856	0.814	0.870
SVM	no assumption	0.893	0.837	0.864
RF	no assumption	0.893	0.824	0.862

4.2 Real data

Li et al. (2020) [28] conducted a meta-analysis that integrated 151 datasets related to COVID-19, encompassing patient symptoms and routine test results. Following a preprocessing phase, 46 clinical variable attributes were selected as explanatory inputs for a classification task. These variables included age, sex, serum levels of neutrophils (both continuous and ordinal), serum levels of leukocytes (both continuous and ordinal), serum levels of lymphocytes (both continuous and ordinal), results from CT scans, results from chest X-rays, reported symptoms (such as diarrhea, fever, coughing, sore throat, nausea, and fatigue), body temperature, and underlying risk factors (including renal diseases and diabetes). Additionally, the class attribute was defined as a binary outcome with levels "H1N1" and "COVID-19." By employing machine learning techniques, the authors reanalyzed the data to explore correlations among the explanatory variables and developed a computational classification model aimed at discriminating between COVID-19 patients and influenza patients based solely on clinical variables.

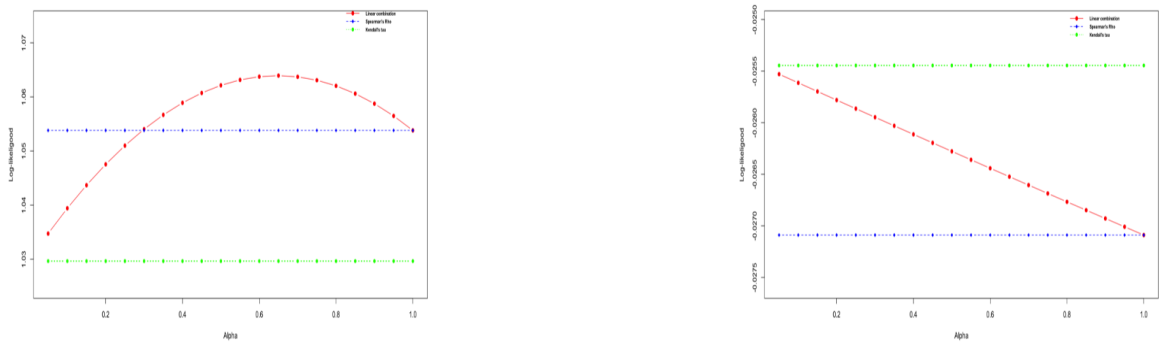


Figure 4: Scatter plot and classification results of Covid-19 data
Panel (a): the association of "age"and "temperate".
Panel (b): the association of "Community Transmission" and "Serum Level White Blood."

Using a copula density with the estimated values of dependence parameters of Spearman's $\hat{\rho}$, Kendall's $\hat{\tau}$ as well as their convex combination, we have tried to classify the class label attribute. We observed that in some cases the convex combination of dependency parameter make a more informative joint density function, see e.g. Figure 4-a for this information about the dependency parameter between "age" and the "temperature". As this figure reveals, the log-likelihood for their joint distribution got the optimum value for the convex combination of $\hat{\rho}$ and $\hat{\tau}$ with $\alpha = 0.62$. Another surprising result is that in some situations, utilizing Kendall's $\hat{\tau}$ as the dependency parameter results in superior outcomes compared to $\hat{\rho}$, particularly when the attributes are nominal in scale. This is exemplified in Figure 4-b, which shows the behavior of the log-likelihood function for the joint distribution of "Community Transmission"

and "Serum Level White Blood." Furthermore, in accordance with algorithm 3, we categorize the patients into two groups: "H1N1" and "COVID-19." The results of this classification are presented in Table 2.

Table 2: Accuracy of classification in Covid-19 dataset

Considered value of parameter	Parameter of dependency	Sensitivity	Specificity	Accuracy
$\hat{\theta} = \hat{\rho}$	Spearman's ρ	0.642	0.658	0.601
$\hat{\theta} = \hat{\tau}$	Kendall's τ	0.701	0.694	0.713
$\hat{\theta} = \hat{\theta}_{cc}$	convex concordance	**0.745**	**0.781**	**0.809**
$\hat{\theta} = 0$	independent attributes	0.657	0.603	0.700

5 Conclusion

A copula-based algorithm is utilized within a classification framework. In this context, the best convex combination of Spearman's rank correlation coefficient ρ and Kendall's association τ is specified to make an accurate classification. Both simulation studies and analyses of real data have demonstrated that the proposed copula-based algorithm can be advantageous when the variables exhibit nonlinear relationships.

The concepts presented in this paper can be further developed in various ways. This approach could be applied to a multi-class classification. Additionally, exploring conditional copula that incorporates the connecting copula given some features would be intriguing. Furthermore, the application of associating copulas for classification tasks, such as support vector machines, discriminant analysis, and naive Bayes classification, presents an area of interest. We aim to expand our findings to longitudinal data, where the outcome variables are interconnected through specific copulas. Ultimately, it may be worthwhile to investigate the results of this study utilizing weak Concordance measures as proposed by as proposed by Liebscher (2014)[14], see also Mesiar et al. (2024)[17], considering, e.g., Spearman's footrule.

Acknowledgement

The work of the second author was supported by the project VEGA 1/0036/23.

References

[1] A. Adadi, *A survey on data-efficient algorithms in big data era*, Journal of Big Data, **8**(1) (2021), 1-54. https://doi.org/10.1186/s40537-021-00419-9

[2] C. Chang, H. Hu, X. Zhao, S. Li, Q. Lin, *Advancing cancer document classification with random forest*, Academic Journal of Science and Technology, **8** (2023), 278-280. https://doi.org/10.54097/ajst.v8i1.14333

[3] Y. Chen, *A copula-based supervised learning classification for continuous and discrete data*, Journal of Data Science, **14**(4) (2016), 769-782. https://doi.org/10.6339/JDS.201610_14(4).0010

[4] F. M. L. Di Lascio, *Coclust: An r package for copula-based cluster analysis*, Recent Applications in Data Clustering, **93** (2018).

[5] A. Dolati, M. Úbeda-Flores, *On measures of multivariate concordance*, Journal of Probability and Statistical Science, **4**(2) (2006), 147-163. https://doi.org/10.1007/s00357-018-9263-4

[6] H. Dong, X. Xu, H. Sui, F. Xu, J. Liu, *Copula-based joint statistical model for polarimetric features and its application in polsar image classification*, IEEE Transactions on Geoscience and Remote Sensing, **55**(10) (2017), 5777-5789. https://doi.org/10.1109/TGRS.2017.2702743

[7] M. Esfahani, M. Amini, G. R. Mohtashami-Borzadaran, A. Dolati, *A new copula-based bivariate Gompertz–Makeham model and its application to COVID-19 mortality data*, Iranian Journal of Fuzzy Systems, **20**(3) (2023), 159-177. https://doi.org/10.22111/ijfs.2023.7645

[8] N. Hammami, M. Bedda, F. Nadir, *Probabilistic classification based on copula for speech recognitation, an overview*, In 2013 International Conference on Computer Applications Technology (ICCAT), (2013), 1-3. https://doi.org/10.1109/ICCAT.2013.6700760

[9] F. Han, H. Liu, *Semiparametric principal component analysis*, In Advances in Neural Information Processing Systems, (2012), 171-179. https://proceedings.neurips.cc/paper/2012/file/4d67b0a97c1668247aa0b0b66f1c7b3d-Paper.pdf

[10] F. Han, H. Liu, *High dimensional semiparametric scale-invariant principal component analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **36**(10) (2014), 2016-2032. https://doi.org/10.1109/TPAMI.2014.2321374

[11] K. Jajuga, D. Papla, *Copula functions in model based clustering*, In From Data and Information Analysis to Knowledge Engineering, (2006), 606-613. https://doi.org/10.1007/3-540-33094-0_60

[12] W. Kruskal, *Ordinal measures of association*, Journal of the American Statistical Association, **53**(284) (1958), 814-861. https://doi.org/10.1080/01621459.1958.10501481

[13] S. Lall, D. Sinha, A. Ghosh, D. Sengupta, S. Bandyopadhyay, *Stable feature selection using copula-based mutual information*, Pattern Recognition, **112** (2020), 107697. https://doi.org/10.1016/j.patcog.2020.107697

[14] E. Liebscher, *Copula-based dependence measures*, Dependence Modeling, **3**(2) (2014). https://doi.org/10.1515/dem-2014-0009

[15] D. Meriem, I. Amrani, T. Balenghien, A. Bennouna, M. Abik, *A procedure to improve binary classification models and categorize features: The case of the distribution of three mosquito species in Morocco*, Revue D'Intelligence Artificielle, **38**(4) (2024). https://doi.org/10.18280/ria.380402

[16] R. Mesiar, A. Kolesárová, A. Sheikhi, *Convex concordance measures*, Fuzzy Sets and Systems, **441** (2022), 366-377. https://doi.org/10.1016/j.fss.2022.03.012

[17] R. Mesiar, A. Kolesárová, A. Sheikhi, S. Shvydka, *Convex weak concordance measures and their constructions*, Fuzzy Sets and Systems, **478** (2024), 108841. https://doi.org/10.1016/j.fss.2024.108841

[18] R. Mesiar, A. Sheikhi, *Nonlinear random forest classification, a copula-based approach*, Applied Sciences, **11**(15) (2021), 7140. https://doi.org/10.3390/app11157140

[19] R. B. Nelsen, *An introduction to copulas*, Springer Science and Business Media, 2006. https://doi.org/10.1007/0-387-28678-0

[20] O. Ozdemir, T. G. Allen, S. Choi, T. Wimalajeewa, P. K. Varshney, *Copula-based classifier fusion under statistical dependence*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **40**(11) (2017), 2740-2748. https://doi.org/10.1109/TPAMI.2017.2673146

[21] J. Rahmanishamsi, A. Dolati, M. R. Aghabozorgi, *A copula-based ica algorithm and its application to time series clustering*, Journal of Classification, **35**(2) (2018), 230-249. https://doi.org/10.1007/s00357-018-9263-4

[22] C. Rung-Ching, W. E. Manongga, C. Dewi, *Recursive feature elimination for improving learning points on hand-sign recognition*, Future Internet, **14**(12) (2022). https://doi.org/10.3390/fi14120584

[23] R. Salinas-Gutiérrez, Q. Angélica Hernández, C. P. Abraham Montoya, H. C. Alberto López, J. M. Marquez Romero, *MLCOPULA: Classification models with copula functions*, CRAN package, 2025. https://cran.r-project.org/package=MLCOPULA

[24] R. Salinas-Gutiérrez, A. Hernández-Aguirre, M. J. Rivera-Meraz, E. R. Villa-Diharce, *Using gaussian copulas in supervised probabilistic classification*, In Soft Computing for Intelligent Control and Mobile Robotics, (2010), 355-372. https://doi.org/10.1007/978-3-642-11809-4_28

[25] R. Salinas-Gutiérrez, A. Hernández-Aguirre, E. Villa-Diharce, *Copula selection for graphical models in continuous Estimation of Distribution Algorithms*, Computational Statistics, **29** (2014), 685-713. https://doi.org/10.1007/s00180-014-0533-4

[26] M. Scarsini, *On measures of concordance*, Stochastica, **8** (1984), 201-218.

[27] A. Sklar, *Fonctions de repartition en dimensions et leurs marges*, Publications de l'Institut de statistique de l'Université de Paris, **8** (1959), 229-231.

[28] W. Tse Li, J. Ma, N. Shende, G. Castaneda, J. Chakladar, J. C. Tsai, L. O. Honda, J. Xu, L. M. Wong, et al., *Using machine learning of clinical data to diagnose covid-19: A systematic review and meta-analysis*, BMC Medical Informatics and Decision Making, **20**(1) (2020), 1-13. https://doi.org/10.1186/s12911-020-01134-5

[29] A. Voisin, V. A. Krylov, G. Moser, S. B. Serpico, J. Zerubia, *Supervised classification of multisensor and multiresolution remote sensing images with a hierarchical copula-based approach*, IEEE Transactions on Geoscience and Remote Sensing, **52**(6) (2013), 3346-3358. https://doi.org/10.1109/TGRS.2013.2242691

[30] T. T. Wong, N. Y. Yang, G. H. Chen, *Hybrid classification algorithms based on instance filtering*, Information Sciences, **520** (2020), 445-455. https://doi.org/10.1016/j.ins.2019.10.028

[31] H. Yue, Q. Meng, D. Chen, *A nonlinear classification model-based method for voltage sag source location in distribution networks with distributed generation*, Electric Power Systems Research, **236** (2024), 110970. https://doi.org/10.1016/j.epsr.2024.110970

[32] S. Zhigang, W. Guotao, L. Pengfei, W. Hui, Z. Min, L. Xiaowen, *An improved random forest based on the classification accuracy and correlation measurement of decision trees*, Expert Systems With Applications, **237** (2024), 121549. https://doi.org/10.1016/j.eswa.2023.121549