UNIVERSIDAD AUTONOMA
DE AGUASCALIENTES

CENTRO DE CIENCIAS BÁSICAS

TESIS

SUPERVISED CLASSIFICATION BASED ON COPULA FUNCTIONS

PRESENTA

Ángela Paulina Pérez Díaz

PARA OBTENER EL GRADO DE MAESTRÍA EN CIENCIAS
CON OPCIÓN A LA COMPUTACIÓN

TUTOR
Dr. Rogelio Salinas Gutiérrez

INTEGRANTES DEL COMITÉ TUTORAL
Dra. Angélica Hernández Quintero
Dr. Oscar Dalmau Cedeño

Aguascalientes, Ags., 19 de febrero de 2018

UNIVERSIDAD AUTONOMA
DE AGUASCALIENTES

M. EN C. JOSÉ DE JESÚS RUÍZ GALLEGOS
DECANO DEL CENTRO DE CIENCIAS BÁSICAS
P R E S E N T E

Por este medio como comité tutoral y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia, nos permitimos emitir **VOTO APROBATORIO** para que la estudiante **ÁNGELA PAULINA PÉREZ DÍAZ** con ID **215825** pueda proceder con la impresión de la tesis titulada **CLASIFICACIÓN SUPERVISADA BASADA EN FUNCIONES CÓPULA**. Es importante mencionar que la tesis ha sido escrita en inglés con el título **SUPERVISED CLASSIFICATION BASED ON COPULA FUNCTIONS**.

Así mismo, manifestamos nuestra aprobación para que la alumna ÁNGELA PAULINA pueda continuar con el procedimiento administrativo para la obtención del grado.

Ponemos lo anterior a su digna consideración y sin otro particular por el momento, le enviamos un cordial saludo.

A T E N T A M E N T E
"Se Lumen Proferre"
Aguascalientes, Ags., a 21 de febrero de 2018.

Dr. Rogelio Salinas Gutiérrez
Tutor de tesis

Dra. Angélica Hernández Quintero
Asesora de tesis

Dr. Oscar Susano Dalmau Cedeño
Asesor de tesis

c.c.p. Ángela Paulina Pérez Díaz - egresada
c.c.p. Dr. Hermilo Sánchez Cruz - Secretario Técnico de la Maestría en Ciencias con opciones a la Computación, Matemáticas Aplicadas.
c.c.p. Archivo

UNIVERSIDAD AUTONOMA DE AGUASCALIENTES

ÁNGELA PAULINA PÉREZ DÍAZ
MAESTRÍA EN CIENCIAS CON OPCIÓN A LA COMPUTACIÓN Y MATEMÁTICAS APLICADAS
P R E S E N T E.

Estimada alumna:

Por medio de este conducto me permito comunicar a Usted que habiendo recibido los votos aprobatorios de los revisores de su trabajo de tesis y/o caso práctico titulado: **"SUPERVISED CLASSIFICATION BASED ON COPULA FUNCTIONS"**, hago de su conocimiento que puede imprimir dicho documento y continuar con los trámites para la presentación de su examen de grado.

Sin otro particular me permito saludarle muy afectuosamente.

A T E N T A M E N T E
Aguascalientes, Ags., a 23 de febrero de 2018
"Se lumen proferre"
EL DECANO

M. en C. JOSÉ DE JESÚS RUÍZ GALLEGOS

c.c.p.- Archivo.

AGRADECIMIENTOS

Al CONACYT por el apoyo brindado para la realización del estudio del posgrado mediante la beca No. 628293.

A la Universidad Autónoma de Aguascalientes donde pude realizar mis estudios e investigación del presente proyecto.

Al Dr. Rogelio Salinas Gutiérrez por su supervisión y dirección en el desarrollo de la investigación, por transmitirme un poco de su gran conocimiento y aceptar ser mi tutor.

A la Dra. Angélica Hernández Quintero por siempre ayudarme, apoyarme cuando lo necesitaba, por su guía y consejos en estos dos años.

Al Dr. Oscar Dalmau Cedeño por su asesoría, sus visitas e invitaciones a Guanajuato que contribuyeron tanto al desarrollo de esta investigación.

Al Dr. Ángel Eduardo Muñoz Zavala y la M. en C. Claudia Nallely Sánchez Gómez por brindarme algo de su tiempo para leer acerca de esta investigación y aceptar ser sinodales en mi examen de grado.

A mi familia, especialmente a mi mamá, Juan, Milo y Pilo que siempre están para mí.

DEDICATORIA

A mi abuelita por su apoyo incondicional.

A mi mamá que cree y confía en mí siempre.

A Pilo que me anima y pone de buenas con sólo existir.

A Juan y Milo, que están conmigo en las buenas y malas.

Contents

List of Tables

List of Figures

Resumen

En esta investigación se muestran los efectos de incorporar funciones cópula a la clasificación supervisada. Las dependencias entre variables en un conjunto de datos pueden ser lineales o no lineales y las funciones cópula son capaces de modelar diferentes tipos de dependencias por lo que resultan una herramienta flexible para modelar dichas dependencias.

La clasificación supervisada es la que se da cuando se tienen clases identificadas así como las características de los objetos que forman parte de cada clase, de esa manera si se tienen nuevos objetos de los cuales se conocen sus características pero no la clase, es posible asignar el objeto a la clase con la que tenga más similitudes según sus características.

Por medio de una serie de experimentos que se realizaron con dos bases de datos diferentes: Base de datos de imágenes de Microsoft [26] y Conjunto de datos de digitos escritos a mano alzada [15], se exponen los resultados que arroja modelar dependencias en clasificación supervisada entre los pares de variables con dependencias más importantes identificados por medio de un modelo gráfico de cadena. Estos resultados son comparados con un clasificador que no considera las dependencias entre variables también conocido como clasificador ingenuo de Bayes.

Los resultados de las muestras analizadas revelan una mejora de hasta 8.3% en algunos casos de clasificación supervisada cuando se incorporan funciones cópula debido a que el desempeño del clasificador es más alto cuando se consideran dependencias que cuando no se hace. Este porcentaje representa una diferencia estadísticamente significativa.

Abstract

This research shows the effects of incorporating copula functions into supervised classification. The dependencies among variables in a dataset can be linear or non-linear and copula functions are able to model different types of dependencies, which make them a flexible tool for modeling these dependencies.

Supervised classification is the one that occurs when the classes are known as well as the characteristics of the objects that are part of each class, that way if there are new objects of which their characteristics are known but not the class, the object can be assigned to the class with which it has most similarities according to its characteristics.

By means of a series of experiments that were carried out with two different databases: Images database from Microsoft [26] and Handwritten digits dataset [15], the results of modeling dependencies in supervised classification between the pairs of variables with more important dependencies identified by means of a graphic chain model are presented. These results are compared with a classifier that does not consider dependencies between variables also known as naive Bayes' classifier.

The results of the analyzed samples reveal an improvement on the average of the experiments of up to 8.3% in some cases of supervised classification when copula functions are incorporated due to the performance of the classifier which is higher when dependencies are considered than when they are not. This percentage represents a statistically significant difference.

1. Introduction

Classification is commonly used nowadays in several sectors like industry [20] and healthcare [1], among others. For instance, in a bank when it is necessary to determine if a person is suitable to receive a bank credit or in a juice company to decide if a fruit is appropriate to create a product, it is necessary to classify.

The terminology in classification can vary, the attributes of an object may also be called features, characteristics, variables, among others [8], the classes can also be called categories or groups.

By knowing the features of certain objects and the class or group where they belong, it is desirable that having a new object whose characteristics are known but not the class to which it belongs, it could be known to which group it should be assigned. There is a wide variety of these characteristics, they could be quality features, requirements or any feature that distinguishes an object from other objects and provides important information about it. Moreover, depending on the type of object and the characteristics that are known, not all the characteristics are always useful for the classification.

There are different types of classification, the one described in the previous paragraph and which is being used in this work is known as supervised classification, whose main objective is to group similar objects into different categories based on their features. In supervised classification, the classes and the features of the objects that are part of those classes are known in advance. This information known as training data provides the classifier with important information to identify the test objects which are the ones to be classified and whose category is unkown.

On the other hand, the use of copula functions is increasing considerably in machine learning thanks to the fact that they provide a flexible tool to build multivariate distributions from marginals which are modeled by different distributions, and a copula function that links these marginals [7].

A copula function is a probability distribution whose main objective is to model dependencies among variables. Copula functions join multivariate distribution functions to their one-dimensional marginal distribution functions [19].

Copula theory, introduced by Sklar in [33] to separate the effect of dependence from the effect of marginal distribution in a joint distribution, allows modeling linear and non-linear dependencies [30].

The intention in this work is to take into account, when classifying, not only the features of the objects but also the most important dependencies among them, to use a graphical model as a tool to identify the most important dependencies and avoid the computational cost that can result from taking all the dependencies. It is intended to observe an improvement in the performance of classifiers when modeling dependencies by using gaussian kernel densities, a graphical model and copula functions.

This work is presented in 7 chapters, the Chapter 2 presents the state of the art where the main concepts to understand the methodology used are explained in detail and the literature review is presented, in that chapter an analysis of the proposals of various authors with researches related with this study is done.

Chapter 3 addresses the research problem where the objectives and hypothesis are raised, the motivation an research question are also presented.

The methodology of the research and how the investigation was carried out is presented in Chapter 4.

Chapter 5 presents the experiments performed and the results obtained, it also describes the databases used.

Chapter 6 presents the discussion, in that chapter an analysis of the results obtained as well as the limitations of the study, recommendations and future work that could continue after this research is performed.

Finally, in Chapter 7, the conclusions are presented.

2. State of the Art

2.1 Classification

A class is a group of similar objects but not necessarily identical, they could be associated by their similar features but they are different from objects of other classes. The main idea on classification is to identify features from objects and assign the objects to the class where they fit best by taking into account the feature values.

According to [36], classification is the grouping together of similar objects. If each object is characterized by d variables, classification can be performed according to rational criteria. Depending on the criteria used, an object could potentially belong to several classes.

Humans are able to naturally classify since the recognition of objects is done instinctively on a daily basis, as well as the distinguishment of different sounds, smells and, visually, the recognition of many things like faces and objects. It is even possible to distinguish one object from another when they are in motion, with different shapes, lighting, etc. Because of the ease with which humans recognize objects, it is hard to transmit the process to do it to a machine [9].

Classification is usually the final step in a process [9] which is briefly outlined:

I The first step is sensing, where data is obtained, this could be done for instance with a camera to obtain images, from a microphone to obtain sounds, etc. The information obtained should be quality data to get the necessary attributes.

II The next step is pre-processing, in this step the object is conditioned for segmentation, an image can be smoothed or the noise of an image, video, audio can be removed.

III Segmentation, this step is where the object is partitioned into regions of interest, an image can be segmented in background and foreground, having in foreground the features of interest and in background everything else.

IV Post-processing is the step that prepares the object to the feature extraction, for example separated objects can be joint, holes can be filled, etc.

V The next step is feature extraction, in this stage, the attributes of the objects whose values must be similar for objects in a particular class and different from the values of the rest of the classes, are identified. The features selected for classification must be meaningful and provide significant information depending on the application of the classification. From the same object, different attributes can be used if the need for classification changes, for instance if a group of people is classified by gender the features extracted would be different from one where they are classified by the language they speak.

VI In the stage of classification, the objects are assigned to certain categories or classes based on their feature information.

Classification can be either supervised or unsupervised. Supervised classification is used when the classes and the attributes of some objects that belong to the classes are known in advance thanks to some training information; on the other hand if the classes are unknown and a set of objects is provided to a program in order for it to identify classes by recognizing similar features in a subset of objects and different from the rest of subsets, then it is unsupervised classification also known as clustering [9]. For this work, supervised classification is used.

2.1.1 Supervised Classification

Supervised learning or classification is a process that occurs when a set of rules are learned from sample objects known as training data and an algorithm is implemented to identify those rules in new samples known as test data [9].

When using email it is often easily identifiable if the received mail is spam or not, however it is not optimal nor convenient to identify spam by our own so, this task is delegated to automated filters that move or delete these kind of emails.

"An email filter is based on a set of rules applied to each incoming message, tagging it as spam or "ham" (not spam). Such a filter is an example of a supervised classification algorithm." [8], these rules can be certain words on the subject or body of the email, the extension of some files, etc., that identify an email as spam so, when a new message arrives if it fits with the rules learned by the filter, it will classify it as spam.

An effective filter as well as any classifier should be able to successfully identify the greatest amount of spam and unwanted emails without loosing legit emails which means that classification must be accurate and efficient enough to not become a bottleneck [8].

There are multiple methods of supervised classification, some of them based on probabilistic classification and some others based on non probabilistic classification. Some of the most known methods are briefly explained as follows.

According with [39] a support vector machine tries to identify support vectors which are observations that are in the edge of an area that represents a limit between these classes of observations and other classes of observations. The space between those regions is known as the margin between classes.

Neural networks are learning algorithms based on the human brain, each neuron is connected to other thousands of neurons and communicates with them. Learning is achieved by adjusting the weights on the connections between nodes, which are analogous to synapses and neurons [11].

A decision tree is a tree where each internal node is associated to a decision and the leaf nodes are usually associated with a result or class. Every internal node tests one or more feature values that lead two or more links. Each link is associated with a possible value of the decision [18].

Nearest neighbor classification [10] is an automatic machine learning method whose main objective is to label unseen query objects until two or more classes are identified. As any classifier, it requires training data with labels, thus is a supervised classification instance. In the simplest variant, the query object inherits the label from the closest sample object in the training set. Common variants extend the decision set from the single nearest neighbor within the training data to the set of k nearest neighbors for any $k > 1$.

2.1.2 Probabilistic Classification

Probabilistic classification uses probabilistic distributions to assign the objects to a class, like multivariate normal classifiers, bayesian network classifiers and even classifiers based on copula functions among others [22]. However, there also exist non probabilistic classifiers which are based on different methods that exclude the use of probability on them like neural networks or support vector machines. For this work, a probabilistic classifier is used since copula functions are part of the classification process executed.

2.2 Maximum Likelihood

According with [2], "The joint density function of n random variables $X_1, ..., X_n$ evaluated at $x_1, ..., x_n$, say $f(x_1, ..., x_n; \theta)$, is referred to as a likelihood function. For fixed $x_1, ..., x_n$ the likelihood function is a function of θ and often is denoted by $L(\theta)$.

If $X_1, ..., X_n$ represents a random sample from $f(x; \theta)$, then:

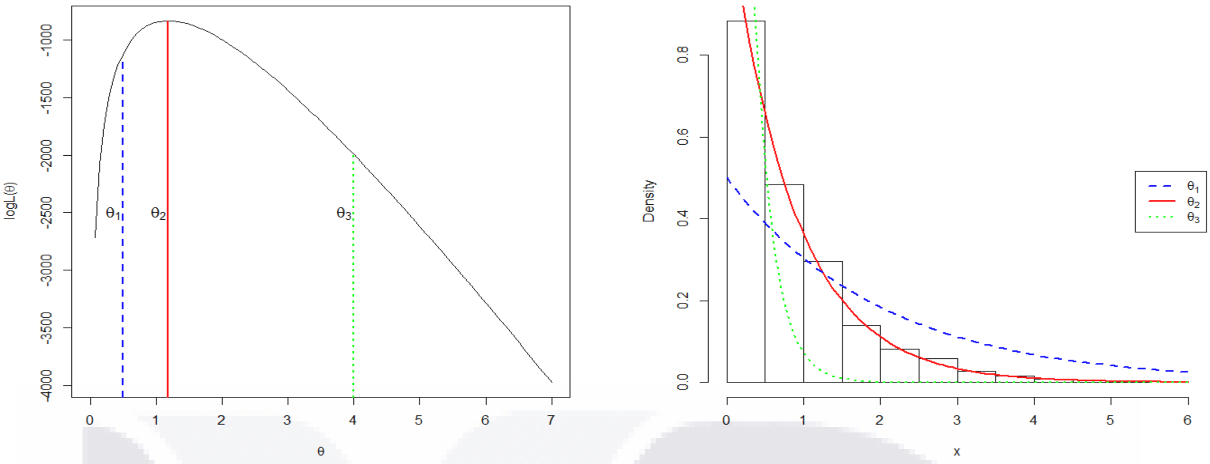
$$L(\theta) = f(x_1; \theta) \cdots f(x_n; \theta), \tag{2.1}$$

Let $L(\theta) = f(x_1, ..., x_n; \theta), \theta \in \Omega$, be the joint pdf of $X_1, ..., X_n$. For a given set of observations, $(x_1, ..., x_n)$, a value $\hat{\theta}$ in Ω at which $L(\theta)$ is a maximum is called a maximum likelihood estimate (MLE) of θ. That is, $\hat{\theta}$ is a value of θ that satisfies: "

$$\begin{gathered} f(x_1, ..., x_n; \hat{\theta}) = \max f(x_1, ..., x_n; \theta) \\ \theta \in \Omega \end{gathered} \tag{2.2}$$

An example of the use of maximum likelihood can be seen in the graphs shown next. In Figure 2.1 (a) The log-likelihood function of a sample of random exponential values is shown, where θ_2 is the parameter that maximizes the function while θ_1 and θ_3 are some other parameters. Maximizing the likelihood is equivalent to maximizing the log-likelihood [4].

In Figure 2.1 (b), θ_1, θ_2 and θ_3 have been represented in the histogram of the same exponential density sample and it can be noticed that θ_2, the parameter that maximizes the log-likelihood function, is the one that fits the most to the model. Therefore θ_2 becomes the maximum likelihood estimate.



(a) (b)

Figure 2.1: (a) Log-likelihood function of an exponential density sample. (b) Histogram of the same exponential density sample with three different proposals for the parameter θ represented.

2.3 Copula Theory

Copula theory was introduced by Sklar [33] to separate the effect of dependence from the effect of marginal distributions in a joint distribution, the use of copula functions in this work is to model dependencies. The separation between marginal distributions and a dependence structure provides flexibility even when the marginals are not of the same type.

Definition 1 *A copula function is a joint distribution function of standard uniform random variables. That is,*

$$C(u_1, u_2, \ldots, u_d) = P[U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d],$$

where $U_i \sim U(0,1)$ *for* $i = 1, 2, \ldots, d$.

Theorem 1 (Sklar's Theorem) *Let* F *be a* d*-dimensional distribution function with marginals* $F_1, F_2, \ldots, F_d$*, then there exists a copula* C *such that for all* x *in* $\overline{\mathbb{R}}^d$ *,*

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)),$$

where $\overline{\mathbb{R}}$ *denotes the extended real line* $[-\infty, \infty]$*. If* $F_1(x_1), F_2(x_2), \ldots, F_d(x_d)$ *are all continuous,* *then C is unique. Otherwise, C is uniquely determined on $Ran(F_1) \times Ran(F_2) \times \ldots \times Ran(F_d)$, where Ran stands for the range.*

Due to Sklar's theorem, any d-dimensional density can be represented as:

$$f(x_1, x_2, \ldots, x_d) = c(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \times \prod_{i=1}^{d} f_i(x_i), \tag{2.3}$$

where c is the density of the copula C, $F_i(x_i)$ is the marginal distribution function of random variable X_i, and $f_i(x_i)$ is the marginal density of variable X_i. Equation (2.3) shows that the dependence structure is modeled by the copula function.

For this research, two-dimensional parametric copula functions are used to model the dependence structure of random variables associated by a joint distribution function. Table 2.1 shows the distribution functions of the following bivariate copula functions which are used in this investigation: Independent, Ali-Mikhail-Haq (AMH), Clayton, Farlie-Gumbel-Morgenstern (FGM), Frank, Gaussian and Gumbel. The dependence parameter θ regulates the strength of association between variables.

Figure 2.2 shows the dependence structure for each copula where θ is a parameter belonging to a subset especified in Tables 2.1 and 2.2 and, as can be seen, the dependence structure is different for every copula. For instance, AMH and FGM model modest dependence. AMH, Clayton, FGM, Frank and Gaussian can model positive and negative dependence but Gumbel copula does not model negative dependence. The gaussian copula models strong dependence on extreme values and weak dependence on centered values, etc. These copula functions are chosen because they cover a wide range of dependencies. The reader interested in copula theory is referred to [19]. The density functions of these copulas are shown in Table 2.2.

The dependence parameter θ of a bivariate copula function can be estimated through the maximum likelihood method (ML). The one-dimensional log-likelihood function, see Equation(2.4), is maximized and its optimal value is used as parameter since it has better properties than other estimators as explained in [38].

$$\ell\left(\theta; \{(u_{1i}, u_{2i})\}_{i=1}^{n}\right) \approx \sum_{i=1}^{n} \log\left(c(u_{1i}, u_{2i}; \theta)\right), \tag{2.4}$$

Assuming that the marginal distributions are known, the observations $\{(u_{1i}, u_{2i})\}_{i=1}^{n}$ in Equation (2.4) are obtained by using the marginal distribution functions of variables X_1 and X_2 [30].

Table 2.1: Bivariate Copula Distributions.

Copula	**Description**
Independent	$C(u_1, u_2) = u_1 u_2$
AMH	$C(u_1, u_2) = \dfrac{u_1 u_2}{1 - \theta(1 - u_1)(1 - u_2)}; \theta \in [-1, 1)$
Clayton	$C(u_1, u_2) = max\left\{\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, 0\right\}; \theta \in [-1, \infty)\backslash\{0\}$
FGM	$C(u_1, u_2) = u_1 u_2 \left(1 + \theta(1 - u_1)(1 - u_2)\right); \theta \in [-1, 1]$
Frank	$C(u_1, u_2) = -\dfrac{1}{\theta} \ln\left(1 + \dfrac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right); \theta \in (-\infty, \infty)\backslash\{0\}$
Gaussian	$C(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \dfrac{e^{-\frac{1}{2}t' \sum^{-1} t}}{2\pi \lvert \sum \rvert^{1/2}} dt_1 dt_2; \theta \in (-1, 1)$ where $\sum$ is a correlation matrix with $\sum_{12} = \theta$
Gumbel	$C(u_1, u_2) = \exp\left(-\left(\tilde{u}_1^{\theta} + \tilde{u}_2^{\theta}\right)^{1/\theta}\right); \theta \in [1, \infty)$ where $\tilde{u}_1 = -\ln(u_1)$ and $\tilde{u}_2 = -\ln(u_2)$

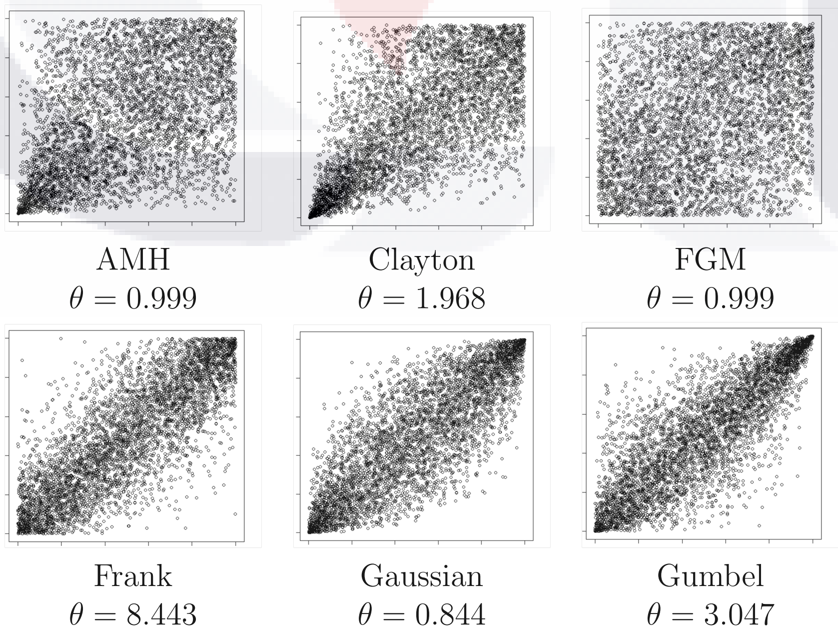


Figure 2.2: Bivariate dependence structure with different copulas.

Table 2.2: Bivariate Copula Densities.

Copula	**Description**
Independent	$c(u_1, u_2) = 1$
AMH	$c(u_1, u_2; \theta) = \dfrac{1 + \theta(u_1 + u_2 + u_1u_2 - 2) - \theta^2(u_1 + u_2 - u_1u_2 - 1)}{(1 - \theta(1 - u_1)(1 - u_2))^3}; \theta \in [-1, 1)$
Clayton	$c(u_1, u_2; \theta) = (1 + \theta)(u_1u_2)^{-\theta-1}\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-2-1/\theta}; \theta \in [-1, \infty)\backslash\{0\}$
FGM	$c(u_1, u_2; \theta) = 1 + \theta(1 - 2u_1)(1 - 2u_2); \theta \in [-1, 1]$
Frank	$c(u_1, u_2; \theta) = \dfrac{-\theta(e^{-\theta} - 1)e^{-\theta(u_1+u_2)}}{((e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1) + (e^{-\theta} - 1))^2}; \theta \in (-\infty, \infty)\backslash\{0\}$
Gaussian	$c(u_1, u_2; \theta) = \left(1 - \theta^2\right)^{-1/2} \exp\left(-\dfrac{(x_1^2 + x_2^2 - 2\theta x_1x_2)}{2(1 - \theta^2)} + \dfrac{(x_1^2 + x_2^2)}{2}\right); \theta \in (-1, 1)$ where $x_1 = \Phi^{-1}(u_1)$ and $x_2 = \Phi^{-1}(u_2)$
Gumbel	$c(u_1, u_2; \theta) = \dfrac{\exp\left(-\left(\tilde{u}_1^\theta + \tilde{u}_2^\theta\right)^{-1/\theta}\right)}{u_1u_2} \dfrac{(\tilde{u}_1\tilde{u}_2)^{\theta-1}}{\left(\tilde{u}_1^\theta + \tilde{u}_2^\theta\right)^{2-1/\theta}} \left((\tilde{u}_1^\theta + \tilde{u}_2^\theta)^{1/\theta} + \theta - 1\right); \theta \in [1, \infty)$ where $\tilde{u}_1 = -\ln(u_1)$ and $\tilde{u}_2 = -\ln(u_2)$

2.4 Bayes Theorem

As explained before, this work is done through a probabilistic classifier by using Bayes theorem shown in Equation (2.5) [9], which proposes the estimation of conditional probability of an event "A", given event "B" but it is necessary to know in advance the conditional probability of "B" given "A".

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{2.5}$$

For the purpose of this work, it is possible to know the probability that an object has to belong to a group A given some features B because the conditional probability of an object that has certain features B when it does belong to a class A is known in advance due to training data. Based on Bayes theorem, there is the naive Bayes classifier [9], which is based on applying Bayes' theorem but assuming that each feature is independent of any other feature given the class, meaning, it does not take into account the association that may exist between its features, a conditional independence is assumed. An example considering three features (B_1, B_2, B_3) can be seen in Equation (2.6)[22]. Notice that $P(A|(b_1, b_2, b_3))$ is the short representation of $P(A|(B_1 = b_1, B_2 = b_2, B_3 = b_3))$ since capital letters represent random variables and lowercase letters represent a value of the random variables.

$$P(A|(b_1, b_2, b_3)) = \frac{P(b_1|A)P(b_2|A)P(b_3|A)P(A)}{P(b_1, b_2, b_3)} \tag{2.6}$$

which can be represented in a graphical model as shown in Figure 2.3.

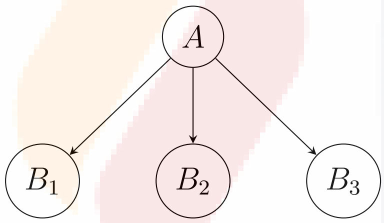


Figure 2.3: Graphical model that represents naive Bayes classifier as seen in Equation (2.6) where conditional independence is assumed.

However there are also classifiers by dependency, as shown in Equation (2.7), that, unlike the previous ones, they consider the association between features of the objects, notice that for Equation (2.7) only three features are considered also [22].

$$P(A|(b_1, b_2, b_3)) = \frac{P((b_1, b_2, b_3)|A)P(A)}{P(b_1, b_2, b_3)} \tag{2.7}$$

which can be represented in a graphical model as shown in Figure 2.4.

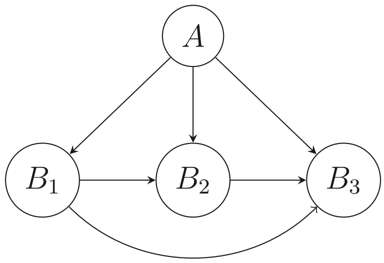


Figure 2.4: Graphical model that represents dependence among variables as seen in Equation (2.7).

2.5 Gaussian Kernel Density

"Kernel Density Estimation estimates the probability density function by imposing a model function on every data point and then adding them together. The function applied to each data point is called a kernel function. For example, a Gaussian function can be imposed on every single data point, making the center of each Gaussian kernel function the data point that it is based on. The standard deviation of the Gaussian kernel function adjusts the dispersion of the function and is called a bandwidth of the function.

Given sufficiently large sample data, KDE can converge to a reasonable estimate of the probability density. As there are no specific finite parameters imposed on the observations, KDE is a nonparametric method. The univariate KDE can be expressed as:

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n}\mathbf{K}\left(\frac{x - X_i}{h}\right), \tag{2.8}$$

where $K(.)$ is the density kernel; x is a test instance point; X_i is a training instance point, which controls the position of the kernel function; h is the bandwidth of the kernel, which controls the dispersion of each kernel; and n is the number of data points" [16].

Because of the flexibility that it provides [13], for this work gaussian kernel density is used to evaluate the marginal distribution, gaussian kernel provides smoothness which is sensible for most datasets. Figure 2.5 shows a representation of kernel density for a sample dataset, the representation is not uni-modal nor symmetric. The normality of data cannot be assumed and that is the reason of using gaussian kernel density in this work.

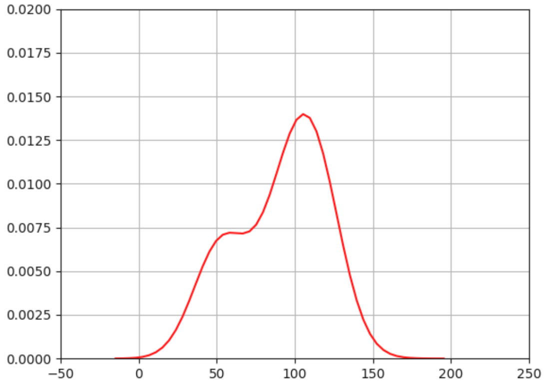


Figure 2.5: Kernel density estimate of a sample dataset.

2.6 Previous Research

Copula functions have been used in finance and other areas [23, 5], they are also being explored in unsupervised classification as can be seen in [6, 17, 27, 35] and in supervised classification [7, 31, 34, 29, 22, 30].

For instance, in [7] the authors solve a classification problem by using vine copulas to model dependencies of multidimensional distributions. In that work, the inversion of Kendall's tau is used to estimate the parameters of bivariate copulas and to model a Dvine copula, the experiments are done using: 1. Product copulas. 2. Only Gaussian copulas and 3. Pair-copulas selected from a catalogue of candidate copulas that consists of Product, Gaussian, Clayton and Gumbel functions.

In [31] four datasets are used to experiment over classification problems with the help of copula theory and suitability of different copula functions in data mining is compared. Eight copula functions are used: Gaussian, Student-t, Clayton, Gumbel, Frank, Hierarchical Clayton, Gumbel and Frank functions whose structure is estimated using Okhrin's algorithm, however no graphical model is employed.

A copula-based bayesian classifier is compared with naive bayes and neural networks classifier in terms of accuracy in [34]. Eight real-world datasets are used however no graphical model is used and the copula parameter is estimated by inversion of pairwise Kendall's tau.

In [29] the Gaussian copula function is used to model probabilistic dependencies in supervised classification for pixel classification in 50 images, by means of accuracy and Tanimoto coefficient, an independent probabilistic model and a copula-based model that takes into account a dependence structure are compared.

In [32] also only Gaussian copula is used for pattern recognition but discrete and mixed features are taken into consideration.

In [22] and [30] copulas AMH, Clayton, FGM, Frank, Gaussian and Gumbel are incorporated to supervised classification. Only a set of experiments with 50 images is done and each copula performance is evaluated separately.

For this work, the use of six copulas is proposed (AMH, Clayton, FGM, Frank, Gaussian and Gumbel) to model dependence, the parameter of dependence is selected with the help of the maximum likelihood method. For the marginal distributions, gaussian kernel densities are used but also a graphical model that takes into account the most important dependencies is employed.

Each copula performance is evaluated separately but also a copula selection is done according to the highest likelihood results. Four sets of experiments are performed, explained in detail in Chapter 5.

3. Research Problem

Supervised classification has been used in various areas of generation and application of knowledge such as industry, healthcare science, computer science, among others. It is well known that is practically impossible to have classifiers be 100% accurate, however, it has consistently been sought to obtain a classifier as precise as possible.

On the other hand, copula functions have been used to help in some areas like economics, finance, civil engineering, data analysis, etc.[3, 21, 23, 28] but little has been explored about copula functions in supervised classification.

Since copulas provide a tool to construct multivariate distributions from marginals of different distributions and a function (copula function) that links those marginals, the intention is to incorporate the advantages of modeling dependencies through copula functions to a supervised classifier.

3.1 Research Question

What would be the performance of a probabilistic classifier if copula functions were incorporated into supervised classification?

3.2 Motivation

Conducting this work responds to the interest of considering if supervised classification with copula functions incorporated is competitive when compared with classifiers that do not use copula functions or do not explicitly take into consideration the dependence among variables.

The improvement in classification by using copula functions could benefit to all those areas that work with supervised classification and can provide a whole new subject to further investigate in pattern recognition.

This research shows the performance of a supervised classifier when copula functions are incorporated and with this, it is intended to spread the results of the work and the future work that it can generate.

3.3 Objective

3.3.1 General Objective

The main objective of this work is to incorporate the copula theory in the design of a probabilistic classifier and evaluate its performance.

3.3.2 Specific Objectives

1. Model dependencies in supervised classification through copula functions.
2. Use a graphical model to select the most important dependencies.
3. Get the performance of a classifier using real datasets and copula functions.
4. Analyze the feasibility of incorporating copula functions into supervised classification.
5. Compare supervised classification with independence vs. dependence.

3.4 Research Hypothesis

The incorporation of copula functions to model dependencies has been successfully applied to some areas, that is why it is believed that incorporating copula functions in supervised classification can improve the results. Given that copulas model dependencies, it is possible to provide more information to the classifier rather than working with independence. Since in this investigation supervised classification is used, modeling dependencies implies getting more training information to work with and having a better overview when the decision of assigning the object to the class is made.

To incorporate the dependency information, a graphical model is used to get enough information to classify without significantly affecting the computational cost.

Since naive Bayes classifier is considered as a good option to classify objects [40], it is believed that using a classifier applying Bayes theorem but considering dependencies, provides a competitive method that can offer alternative solutions to multiple areas.

4. Methodology

4.1 Graphical Chain Model

As known, copula functions can model dependencies among variables. The proposal in this work was to use a graphical model as a tool and select the most important bivariate dependencies. “Graphical models are probability models for multivariate observations to analyze and visualize conditional relationships between random variables encoded by a conditional independence graph” [25].

The dependence structure is based on a chain model which, for a d-dimensional continuous random vector represents a probabilistic model with density [30]:

$$f_{\text{chain}}(\mathbf{x}) = f\left(x_{\alpha_1}\right) \prod_{i=2}^{d} f\left(x_{\alpha_i} | x_{\alpha_{(i-1)}}\right), \tag{4.1}$$

where $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d)$ is a permutation of the integers between 1 and d. An example of a graphical chain model for a three dimensional vector is shown in Figure 4.1 [30].

$$X_{\alpha_1} \longrightarrow X_{\alpha_2} \longrightarrow X_{\alpha_3}$$

$$f_{\text{chain}}(\mathbf{x}) = f\left(x_{\alpha_1}\right) f\left(x_{\alpha_2} | x_{\alpha_1}\right) f\left(x_{\alpha_3} | x_{\alpha_2}\right)$$

Figure 4.1: Joint distribution over 3 variables represented by a graphical chain model.

4.2 Kullback-Leibler Divergence

To properly use a chain model, it is necessary to select the dependencies to be used, one way to select the most important dependencies, the permutation $\vec{\alpha}$, is through Kullback-Leibler divergence (D_{KL}). D_{KL} is an information measure between two distributions [37]. It is always non-negative for any two distributions, and is zero if and only if the distributions are identical [30]. Hence, the Kullback-Leibler divergence can be interpreted as a measure of the dissimilarity between two distributions. Then, the goal is to choose a permutation $\vec{\alpha}$ that minimizes the Kullback-Leibler divergence between the true distribution $f(\mathbf{x})$ of the dataset and the distribution associated to a chain model, $f_{\text{chain}}(\mathbf{x})$. For instance, the Kullback-Leibler divergence between joint densities f and f_{chain} for a continuous random vector $\mathbf{X} = (X_1, X_2, X_3)$ is given by:

$$\begin{aligned} D_{KL}\left(f||f_{\text{chain}}\right) &= E_f\left[\log\frac{f(\mathbf{x})}{f_{\text{chain}}(\mathbf{x})}\right] \\ &= -H(\mathbf{X}) + \int \log\left(f\left(x_{\alpha_1}\right) f\left(x_{\alpha_2}|x_{\alpha_1}\right) f\left(x_{\alpha_3}|x_{\alpha_2}\right)\right) f dx. \end{aligned} \tag{4.2}$$

The first term in Equation (4.2), $H(\mathbf{X})$, is the entropy of the joint distribution $f(\mathbf{x})$ and does not depend on the permutation $\vec{\alpha}$. By using copula theory and Equation (2.3), the second term can be decomposed into the product of marginal distributions and bivariate copula functions.

$$\begin{aligned} D_{KL}\left(f||f_{\text{chain}}\right) &= -H(\mathbf{X}) + \sum_{i=1}^{3} H(X_i) \\ &\quad - \int \log\left(c\left(u_{\alpha_1}, u_{\alpha_2}; \theta_{\alpha_1,\alpha_2}\right)\right) f dx \\ &\quad - \int \log\left(c\left(u_{\alpha_2}, u_{\alpha_3}; \theta_{\alpha_2,\alpha_3}\right)\right) f dx. \end{aligned} \tag{4.3}$$

The second term of Equation (4.3), the sum of marginal entropies, also does not depend on the permutation $\vec{\alpha}$. Therefore, minimizing Equation (4.3) is equivalent to maximize the sum of the last two terms. Once a sample of size n is obtained from the joint density f, the last two terms can be approximated by a Monte Carlo approach:

$$\int \log\left(c\left(u_{\alpha_1}, u_{\alpha_2}; \theta_{\alpha_1,\alpha_2}\right)\right) f dx \approx \frac{1}{n}\sum_{i=1}^{n} \log\left(c\left(u_{1i}, u_{2i}; \theta_{\alpha_1,\alpha_2}\right)\right). \tag{4.4}$$

Through Equation (4.4), the D_{KL} is minimized by maximizing the sum of the log-likelihood for the copula parameters. Note that the log-likelihood allows to estimate the copula parameter and to select the appropriate permutation $\vec{\alpha}$. Finally, by means of copula theory, a graphical chain model for a three dimensional vector has the density

$$f_{\text{chain}}(\mathbf{x}) = f\left(x_{\alpha_1}\right) f\left(x_{\alpha_2}\right) f\left(x_{\alpha_3}\right) c\left(u_{\alpha_1}, u_{\alpha_2}; \hat{\theta}_{\alpha_1,\alpha_2}\right) c\left(u_{\alpha_2}, u_{\alpha_3}; \hat{\theta}_{\alpha_2,\alpha_3}\right) \tag{4.5}$$

And it is represented as Figure 4.2

$$X_{\alpha_1} \xrightarrow{\text{copula}} X_{\alpha_2} \xrightarrow{\text{copula}} X_{\alpha_3}$$

Figure 4.2: Graphical representation of Equation (4.5).

4.3 Evaluation Metrics

To evaluate the performance of the classifiers three evaluation metrics were used. Accuracy that shows the amount of data correctly classified, sensitivity and specificity (for two classes) that show the amount of data correctly classified for positive class and negative class.

One of the classes will be considered as positive and the another one as negative, thus, true positive is the data correctly classified in positive class, false positive is the data incorrectly classified in positive class, true negative is the data correctly classified in negative class and false negative is the data incorrectly classified in negative class as explained in Figure 4.3.

In the case when there are more than two classes, it is not possible to get sensitivity and specificity.

		Truth	
		Positive	Negative
Model	Positive	tp	fp
	Negative	fn	tn

(a)

$$accuracy = \frac{tp + tn}{tp + fp + fn + tn}$$

$$sensitivity = \frac{tp}{tp + fn}$$

$$specificity = \frac{tn}{tn + fp}$$

(b)

Figure 4.3: (a) Confusion matrix for binary classification, tp stands for true positive, fp is false positive, fn is false negative, and tn is true negative. (b) Definitions for accuracy, sensitivity and specificity used in this work.

4.4 Step by Step Methodology

The classifiers to perform the experiments were developed in python. “Python is a general-purpose programming language. That means that it was designed and developed to write software for a wide variety of disciplines” [14], it is an easy and intuitive language, the environment is almost completely platform-independent and the code written in python is platform independent [14]. Another advantage of python is that it is open source, it is an interpreted language. The steps followed to get the evaluation metrics are the following:

I The first step is to extract the attributes of the objects in training data, meaning the ones that have a class assigned.

II It is obtained gaussian kernel distribution for every attribute in each class as described in the following pseudocode.

Pseudocode for getting gaussian kernel distribution

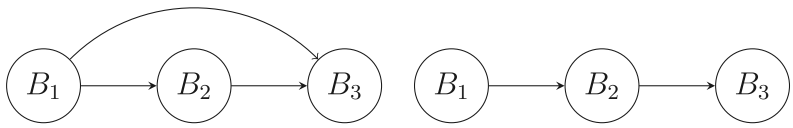
1. for $i = 1$ to $i =$ No. classes
2.   for $j = 1$ to $j =$ No. attributes
3. $y_{ij} =$ probability density function PDF of $class_i$ $attribute_j$
4. $Y_{ij} =$ Integrate y_{ij} to get the cumulative distribution function CDF

III Get the dependence parameter (θ) and log-likelihood for every pair of attributes. To do so, it is necessary the maximum likelihood method, this step is described in the following pseudocode.

Pseudocode for getting θ and maximum log-likelihood

1. for $i = 1$ to $i =$ No. classes
2.   for $j = 1$ to $j =$ No. attributes
3.   for $k = j + 1$ to No. attributes
4.   Maximize (log-likelihood of copula ($class_i$ $attribute_j, class_i$ $attribute_k$))
5.   return maximum log-likelihood, θ that maximizes the log-likelihood

IV Having the values of the dependence parameter and log-likelihood for each pair of attributes, the graphical chain model is formed with the most important dependencies, in the case of three variables instead of taking into account all the dependencies as seen in Figure 4.4 (a), the two most important dependencies were used as seen in Figure 4.4 (b)



(a) (b)

Figure 4.4: (a) Graphical representation of all dependencies between an object with three attributes. (b) Graphical chain model of an object with three attributes.

V Get the copula densities using the extracted parameter for the most important dependencies.

VI Then, using Bayes theorem, it is necessary to get the probability that every test object has of belonging to each class and it is classified where the probability is the highest, the proposed model for a three variables example is shown in Equation 4.6.

$$P(b_1, b_2, b_3|A) = P(b_1|A) \times c(F_{B_1}, F_{B_2}|A) \times P(b_2|A) \times c(F_{B_2}, F_{B_3}|A) \times P(b_3|A) \tag{4.6}$$

VII The final step is to get the evaluation metrics to know the performance of the classification.

5. Experiments and Results

5.1 Images from Microsoft Repository

The first stage of experiments were done with pixel classification. From raster images also known as digital images (the reader interested in knowing what a raster image is, is referred to [24]) taken from [26], having three features: Red (R), Green (G) and Blue (B) and two established classes: background and foreground. The features and the class where the pixels belonged were used, extracted from training data to classify test data. It will be explained with more detail in the next paragraph.

Microsoft repository [26] provided three different images for each picture; see Figure 5.1 the first one is the color image called "ELEPHANT" from where the RGB information is extracted, the second image in gray scale represents the training data for both classes and test data, the third image is correctly classified and it is the image that allowed to evaluate the performance of the implemented classifiers.

(a) (b) (c)

Figure 5.1: (a) Color image. (b) Image that shows the training data for background (dark gray), foreground (white) and test data (gray). (c) Correctly classified image, background (black) and foreground (white).

The experiments were done with 50 images shown in Figure 5.2. Thanks to the color image, three features were obtained: R, G and B. Every feature on each pixel has a value from 0 to 255 that represents the color intensity.

Figure 5.2: 50 color images from Microsoft repository.

A color image can be represented in a 3-dimensional matrix to keep data for R, G and B colors, this is the information that was used as the attributes of every pixel to classify them.
As mentioned in [22], 30 of the 50 images were classified using normal distribution; the classification was made in two ways: without taking into account the association among the features and taking into account the dependencies or association among them. The results obtained in one of the experiments of image called "PERSON6" with a normal distribution and independence between features are shown in Figure 5.3 (b). The evaluation metrics of the classification with normal distribution and independence are: Accuracy - 62.65%, Sensitivity - 37.86% and Specificity - 83.05% [22].
The same image was classified taking into account the dependency among the features, as can be seen in 5.3 (c) and the results were: Accuracy - 82.65%, Sensitivity - 74.83%, Specificity - 89.08% [22].

(a) (b) (c)

Figure 5.3: (a) Correctly classified image. (b) Image classified with normal density and independence between features. (c) Image classified with normal density and dependence between features.

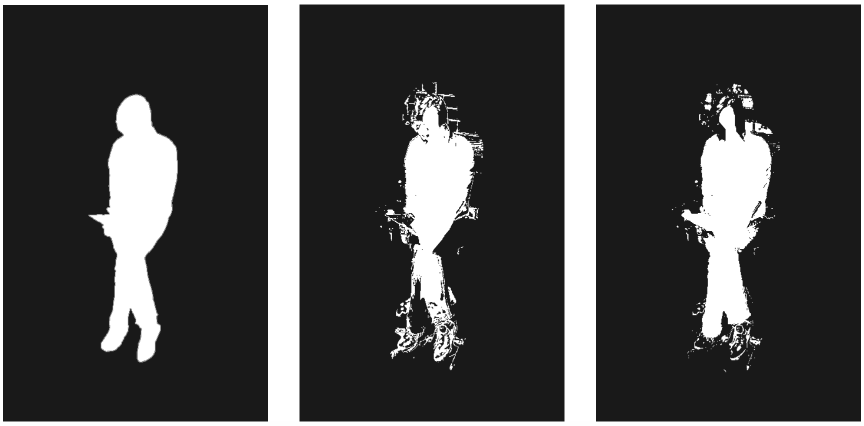
It can be learned from Figure 5.3 that taking into account the dependencies between attributes might provide an improvement in supervised classification, and after experimenting with 30 images a trend was noticed, but since the normality of the information cannot be assumed and because of the flexibility that it offers, it was decided to use the gaussian kernel density instead of the normal density.

As mentioned earlier, one of the advantages of using gaussian kernel density is the flexibility that it provides, this flexible marginal density was used with independence at first, the results are shown in Figure 5.4 (b) of image "PERSON6".

The evaluation metrics for the image shown in Figure 5.4(b) which was classified with gaussian kernel density and independence among features were: Accuracy - 74.98%, Sensitivity - 58.46%, Specificity - 88.77%.

Then, copula functions were incorporated in classifiers with gaussian kernel density; the main objective was to model dependency among the attributes. In order to cover a considerable amount of models, the classification was done with six different density copulas, the ones mentioned before in Table 2.2

The classification was done by getting the copula parameter that maximizes the log-likelihood in the copula function. The 50 images were classified using all six copulas; in Figure 5.4 (c), an image classified with kernel density function and copula Frank is shown. The results on that image were: Accuracy - 82.80%, Sensitivity - 74.03%, Specificity - 90.01%.



(a) (b) (c)

Figure 5.4: (a) Correctly classified image. (b) Image classified with kernel density and by independence. (c) Image classified by Frank copula.

Image 227092, wich is a vase, is another of the 50 images taken from [26] that was classified, in Figure 5.5, the scatterplot in the domain of copulas of this image is shown.

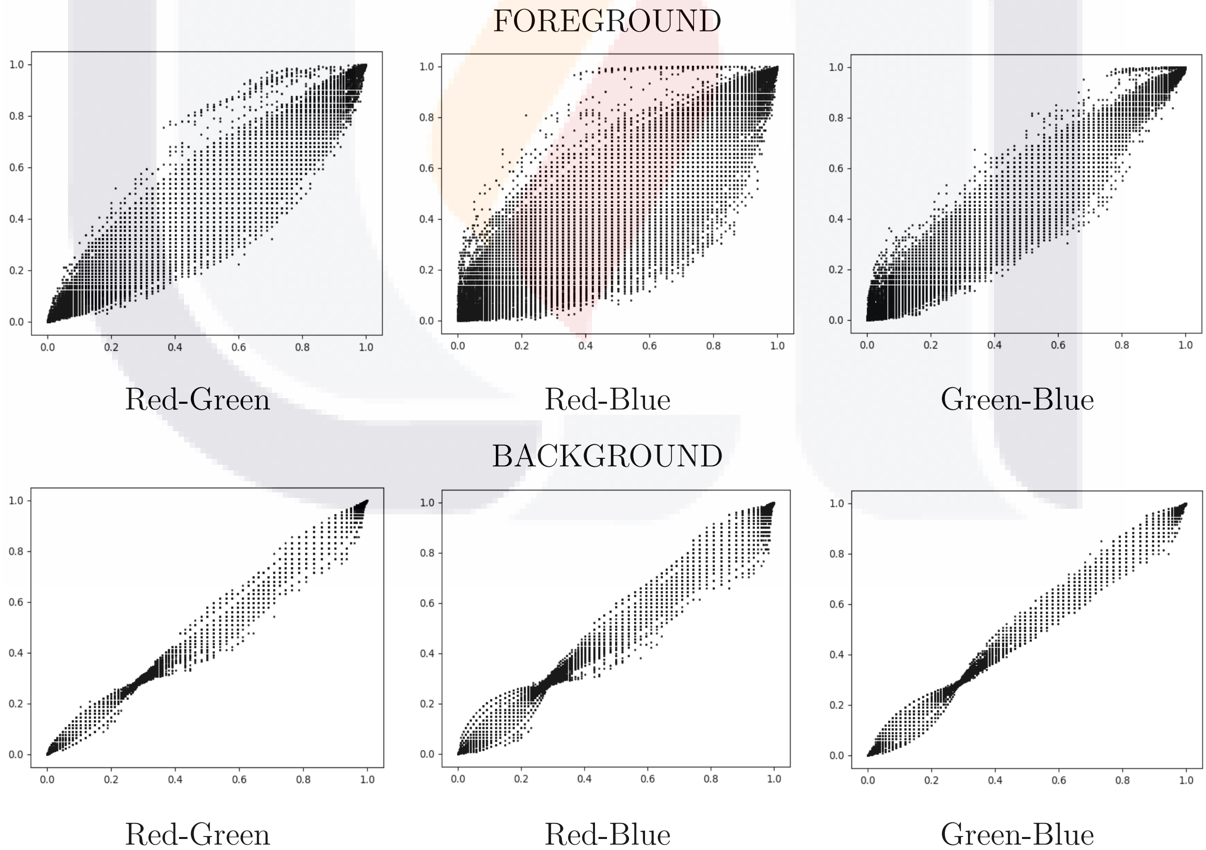
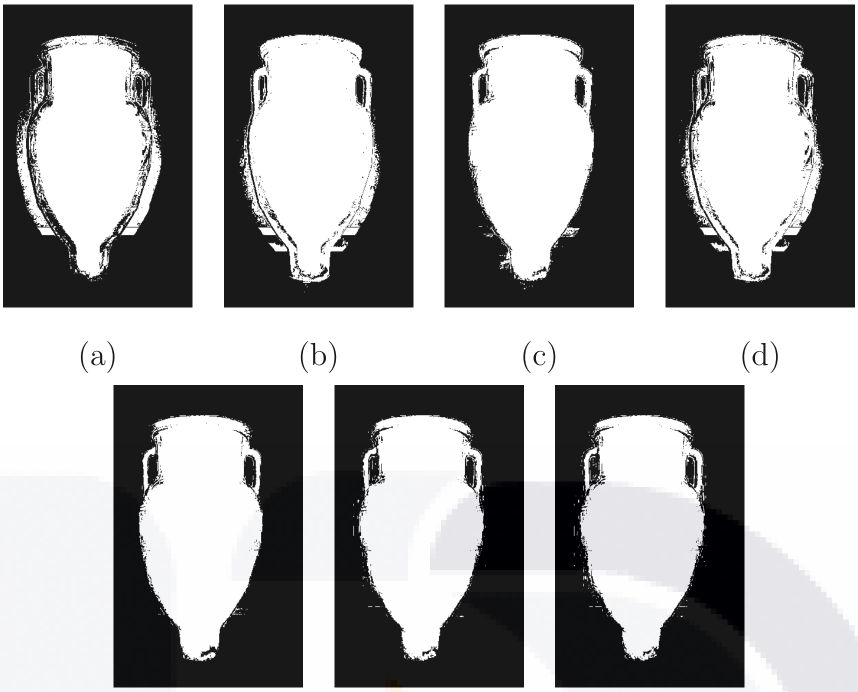


Figure 5.5: Scatterplot of image 227092 in copulas domain (0-1) for each pair of variables.

For the foreground, a dependence similar to an elliptical structure in the three pairs of variables with strong dependence on queues and a weaker dependence on the center can be noticed. On the other hand, for the background, there is a strong dependence near the center of the figure. The image of the vase can be seen in Figure 5.6 where the color, labelling-lasso and corrected classified images are shown. Fig 5.7 shows the image classified with independent, AMH, Clayton, FGM, Frank, Gaussian and Gumbel copulas.

(a) (b) (c)

Figure 5.6: (a) The color image. (b) The labelling-lasso image with the training data for background (dark gray), for foreground (white) and the test data (gray). (c) The correct classification with foreground (white) and background (black).



(a) (b) (c) (d)

(e) (f) (g)

Figure 5.7: (a) Independent copula. (b) AMH copula. (c) Clayton copula. (d) FGM copula. (e) Frank copula. (f) Gaussian copula. (g) Gumbel copula.

The values for the metrics obtained by the classifiers when copulas were incorporated from the 50 images have been summarized in Table 5.1 [22]. All six copulas had a better behaviour than the independent copula which represents no association among features.

An ANOVA test for comparing the mean accuracy among classifiers was done in [30]. The test reports a statistical difference between Clayton, Frank, Gaussian and Gumbel copula functions with respect to the independent copula (p-value < 0.05). The major difference in accuracy with respect to the independent copula is given by the Frank copula. Accuracy, as described in Figure 4.3, shows the amount of pixels that were classified correctly.

Table 5.1: Descriptive results for evaluation metrics, the values with a statistical difference between the copula and the independent are marked with an asterisk.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Independent	79.4	10.8	77.3	16.6	81.3	13.6
AMH	82.9	9.5	80.7	15.9	84.7	11.9
Clayton	86.1*	8.7	83.0	15.7	88.5	9.7
FGM	80.9	9.8	78.9	16.5	82.5	13.2
Frank	87.7*	7.1	87.1	12.2	88.1	9.0
Gaussian	86.0*	10.6	87.1	11.0	85.0	18.6
Gumbel	86.7*	8.2	87.0	10.9	86.5	13.2

5.1.1 Images with Random Sample

In a second stage of experiments the pixels were picked randomly. Instead of having the test data in a frame around the foreground training information as in the previous experiments. Both, the training and test data were randomly selected from the whole image.

The pixels were selected in the following order having 6 sets of pixels as shown in Table 5.2

Table 5.2: Sets of Sample Pixels.

Training data	**Test data**
5%	95%
10%	90%
20%	80%
30%	70%
40%	60%
50%	50%

In Figure 5.8 there are shown the images with 5, 10, 20, 30, 40 and 50 percent of the pixels as training data. The sample was selected randomly for both foreground and background training data. In (a) it can be seen how the image with training and test data looks for 5%, it is a vase but the form is hardly visible because the 95% of the pixels are test pixels. In (b) and (c), the same image but with 10 and 20 percent of the pixels respectively as training data is shown. In Figure 5.8 (d), (e) and (f) the vase figure is more visible since the test pixels are less than in previous images in (a), (b) and (c).

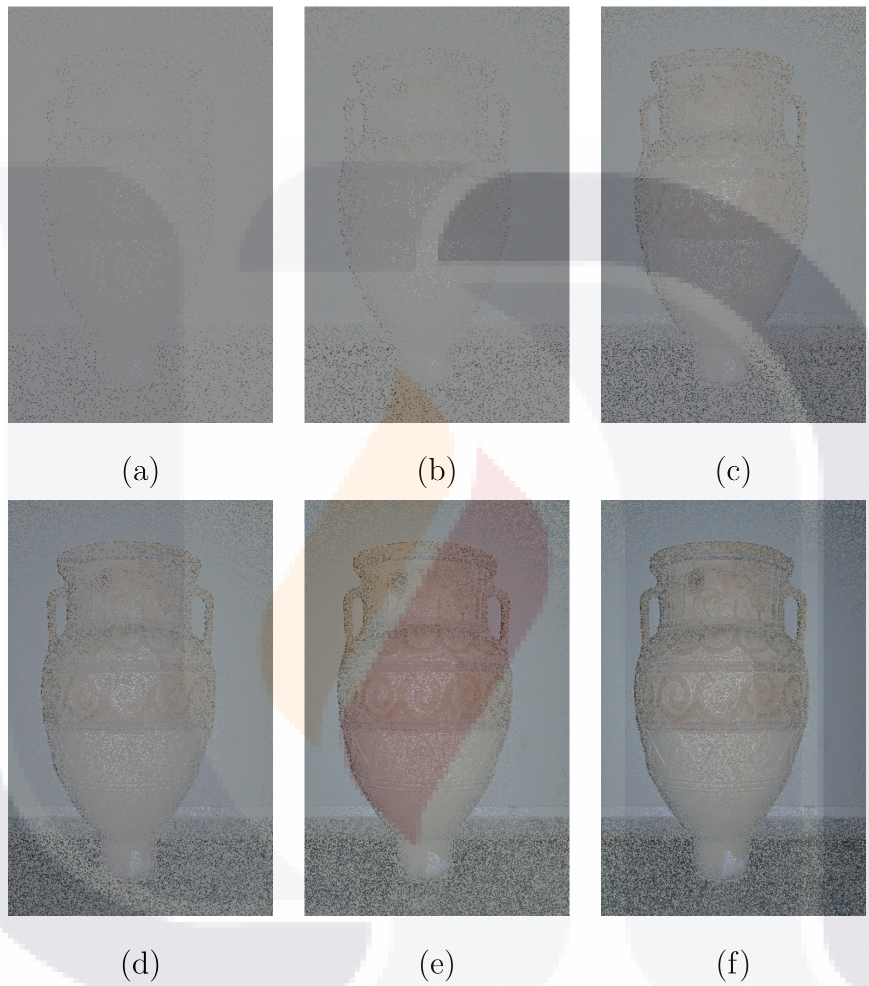


Figure 5.8: (a) Image with 5% of training data. (b) 10%. (c) 20%. (d) 30%. (e) 40%. (f) 50%.

The following Tables (5.3, 5.4, 5.5, 5.6, 5.7 and 5.8) show the mean and standard deviation of the values of metrics (accuracy, sensitivity and specificity) of the 50 images from [26] taking 5%, 10%, 20%, 30%, 40% and 50% of the pixels as training data respectively.

Table 5.3: Descriptive results for evaluation metrics with training data size of 5%, the values with a statistical difference between the copula and the independent are marked with an asterisk.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Independent	84.2	10.5	82.3	14.1	84.5	11.1
AMH	86.6	10.4	83.5	15.3	87.2	11.8
Clayton	90.9*	6.5	89.9	10.2	91.5	6.9
FGM	85.2	10.6	82.7	15.6	85.8	11.9
Frank	91.6*	7.1	92.2	8.8	91.4	7.4
Gaussian	92.4*	6.8	92.9	7.8	92.3	7.5
Gumbel	92.2*	6.7	93.0	7.1	92.0	7.2

Table 5.4: Descriptive results for evaluation metrics with training data size of 10%, the values with a statistical difference between the copula and the independent are marked with an asterisk.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Independent	84.2	10.4	82.3	13.8	84.5	11.0
AMH	86.6	10.4	83.8	15.0	87.1	11.8
Clayton	90.9*	7.4	90.2	9.7	91.2	7.7
FGM	85.3	10.7	83.0	15.2	85.7	11.9
Frank	91.7*	7.0	92.1	8.6	91.5	7.3
Gaussian	92.3*	6.8	92.9	7.8	92.2	7.5
Gumbel	92.1*	6.7	92.9	7.3	92.0	7.2

Table 5.5: Descriptive results for evaluation metrics with training data size of 20%, the values with a statistical difference between the copula and the independent are marked with an asterisk.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Independent	84.2	10.4	82.6	13.6	84.4	11.1
AMH	86.6	10.4	84.1	14.8	87.1	11.8
Clayton	90.9*	6.9	89.6	10.7	91.3	7.1
FGM	85.2	10.8	83.1	15.0	85.7	12.0
Frank	91.6*	7.1	92.3	8.7	91.4	7.5
Gaussian	92.3*	6.9	93.0	7.6	92.1	7.6
Gumbel	92.1*	6.7	93.1	7.1	91.9	7.2

Table 5.6: Descriptive results for evaluation metrics with training data size of 30%, the values with a statistical difference between the copula and the independent are marked with an asterisk.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Independent	84.2	10.4	82.5	13.8	84.4	11.0
AMH	86.6	10.3	84.1	14.8	87.1	11.7
Clayton	91.5*	6.4	90.0	10.7	91.8	6.6
FGM	85.3	10.7	83.1	15.1	85.8	11.9
Frank	91.6*	7.1	92.3	8.6	91.4	7.4
Gaussian	92.3*	6.9	92.9	7.7	92.1	7.6
Gumbel	92.1*	6.8	93.1	7.1	91.8	7.3

Table 5.7: Descriptive results for evaluation metrics with training data size of 40%, the values with a statistical difference between the copula and the independent are marked with an asterisk.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Independent	84.3	10.3	82.5	13.7	85.0	10.1
AMH	86.7	10.3	84.2	14.8	87.1	11.7
Clayton	91.3*	6.5	89.6	10.6	91.7	6.8
FGM	85.4	10.5	83.1	15.1	85.8	11.8
Frank	91.6*	7.0	92.3	8.6	91.4	7.4
Gaussian	92.2*	6.9	92.8	7.7	92.1	7.6
Gumbel	92.1*	6.7	93.1	7.2	91.8	7.2

Table 5.8: Descriptive results for evaluation metrics with training data size of 50%, the values with a statistical difference between the copula and the independent are marked with an asterisk.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Independent	84.3	10.3	82.5	13.6	84.5	10.9
AMH	86.6	10.3	84.1	14.8	87.1	11.7
Clayton	91.2*	6.6	89.2	12.0	91.6	6.8
FGM	85.4	10.5	83.1	15.0	85.9	11.8
Frank	91.6*	7.0	92.3	8.6	91.4	7.4
Gaussian	92.2*	6.9	92.8	7.7	92.1	7.6
Gumbel	92.1*	6.8	93.1	7.2	91.8	7.3

An ANOVA test for comparing the mean accuracy among classifiers was also done for these experiments. The test reports a statistical difference between Clayton, Frank, Gaussian and Gumbel copula functions with respect to the independent copula (p-value < 0.01).

With the results in the mean accuracy for all the experiments with random sample, meaning with 5, 10, 20, 30, 40 and 50 percent of the pixels as training information, it was possible to build the graph shown in Figure 5.9 where it is visible that the mean accuracy of the 50 images is very similar for all different percentages of training data in all seven copulas including independent.

With these results, it is believed that it is feasible to use less training data when using random samples and the results will be equally valid when using more training information.

It can also be seen in Figure 5.9 that the copulas with a statistical difference between them and independent copula (Clayton, Frank, Gaussian and Gumbel) are on top of the graph.

Figure 5.9: The graph shows the accuracy (y axis) of the 50 images with different sample size percentage on training datasets (x axis).

5.1.2 Freehand Images with GIMP Editor

Thinking of the possibility of having users or customers selecting the training and test data, a series of experiments were performed on the images by selecting freehand pixels with the help of GIMP editor, which is a computer program for creating and editing digital images [12] as seen in Figure 5.10.

(a) (b) (c)

Figure 5.10: (a) The color image. (b) Image edited with GIMP, training data for background and foreground (color) and the test data (gray). (c) The correct classification with foreground (white) and background (black).

The experiments were performed in 50 images by following the same procedure. The background and foreground training pixels were selected by freehand, after the training of the classifier, the test pixels (gray in Figure 5.10 (b)) were introduced to the classifier to know the class where they belonged. The results of mean and standard deviation for accuracy, specificity and sensitivity of the 50 images selecting pixels by freehand are shown in Table 5.9.

Table 5.9: Descriptive results for evaluation metrics with GIMP editor, the values with a statistical difference between the copula and the independent are marked with an asterisk.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Independent	83.6	11.4	80.8	15.4	83.9	12.0
AMH	86.5	10.6	82.1	15.8	87.1	11.6
Clayton	91.5*	6.2	87.6	12.8	92.0	6.3
FGM	84.9	11.5	81.3	16.2	85.5	12.5
Frank	92.2*	6.0	90.2	9.4	92.4	6.1
Gaussian	92.2*	7.0	90.9	9.3	92.3	7.4
Gumbel	91.8*	6.7	90.9	9.2	91.8	7.0

5.1.3 Copula Selection

The experiments done so far were using the same copula function for each couple of variables of the most important dependencies, two for the case of the images.

The proposal in the next experiments was to perform a copula selection. Based on the results of log-likelihood for every couple of variables and taking into consideration all six copulas, the algorithm selected the highest values in log-likelihood that could form a graphic chain model. That way, for the example shown in Figure 5.11 where 3 variables are taken into consideration (R,G,B) a copula function can be used for the dependence of the first pair of variables R and G (copula 1) and some other copula function for the dependence of other pair G and B (copula 2). In copula selection, a copula function was used for each couple of variables.

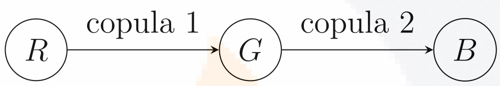


Figure 5.11: Example of a graphic chain model with 3 features (variables).

Table 5.10 show the results of copula selection for all the experiments performed, the asterisk in accuracy represents a statistical difference between independent copula and copula selection.

Table 5.10: Descriptive results for evaluation metrics in all the experiments performed with copula selection, RS stands for Random Sample.

Copula	**Accuracy %**		**Sensitivity %**		**Specificity %**	
Model	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Microsoft database	86.2*	10.6	87.2	11.4	85.3	18.4
Random Sample 5%	92.1*	6.9	93.0	6.8	91.8	7.7
Random Sample 10%	91.8*	7.3	92.8	7.5	91.6	7.8
Random Sample 20%	92.0*	6.7	92.5	8.5	91.9	7.1
Random Sample 30%	92.2*	6.6	92.6	8.7	92.0	6.9
Random Sample 40%	92.0*	6.6	92.0	9.1	92.0	7.0
Random Sample 50%	92.1*	6.6	92.3	8.8	92.0	7.0
Freehand	92.6*	6.4	91.6	8.2	92.7	6.6

Figure 5.12 shows a graph where the mean accuracy of the six copulas, independent copula and

copula selection for the experiments of Microsoft repository, random sample of 5% of training data and freehand are represented in a different line each.

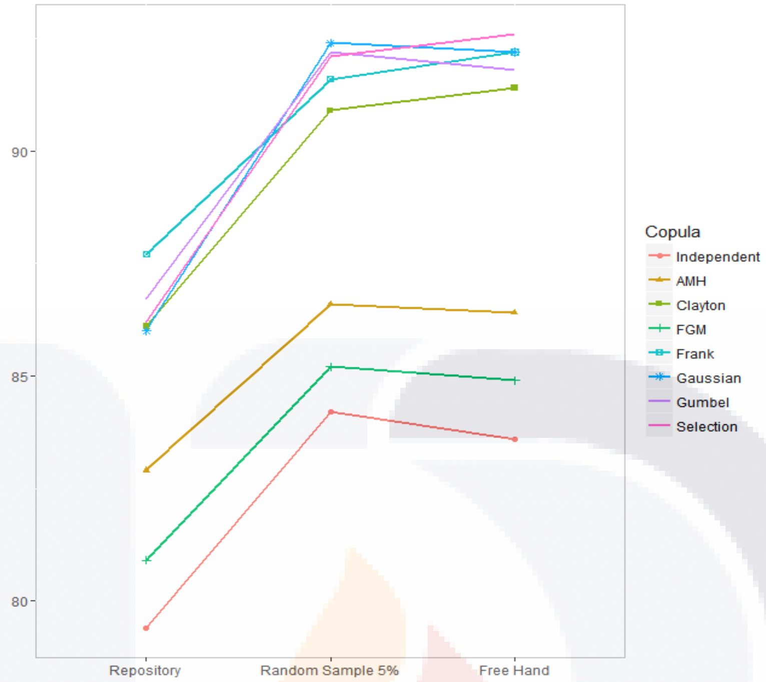


Figure 5.12: The graph shows the accuracy (y axis) of the 50 images with the experiments done from Microsoft repository, random sample of 5% of training data and Freehand (x axis).

5.2 Handwritten Digits Dataset

For this investigation, another database different from the image database [26] was used. This dataset obtained from [15] is a database of 250 samples from 44 writers. The samples written by 30 writers are used for training, and the samples written by the other 14 writers are used for independent testing.

They describe the dataset as follows:

"We create a digit database by collecting 250 samples from 44 writers. The samples written by 30 writers are used for training, cross-validation and writer dependent testing, and the digits written by the other 14 are used for writer independent testing.

We use a WACOM PL-100V pressure sensitive tablet, the tablet sends x and y tablet coordinates and pressure level values of the pen at fixed time intervals (sampling rate) of 100 miliseconds.

These writers are asked to write 250 digits in random order inside boxes of 500 by 500 tablet pixel resolution. Each screen contains five boxes with the digits to be written displayed above.

In our study, we use only (x, y) coordinate information. The stylus pressure level values are ignored" [15].

The objects in this database have not only three features, they have sixteen features and the classes were ten instead of two like the pixels in previous experiments.

For this database, accuracy was the only evaluation metric used. Since the objects are classified in 10 classes the estimation of sensitivity and specificity becomes difficult, as explained in Figure 4.3 sensitivity and specificity are metrics for cases with two classes. The results are the following, shown in Table 5.11.

Table 5.11: Descriptive results for accuracy for Handwritten dataset.

Copula	Accuracy %
Independent	83.6
AMH	87.2
Clayton	87.6
FGM	86.1
Frank	90.8
Gaussian	92.3
Gumbel	90.1
Selection	91.3

It can be noticed that the results in all copulas and copula selection is higher than the result for Independent copula. For this case Gaussian copula got the best result with an improvement of 8.7%.

The information in the dataset was re-distributed randomly to perform 30 new experiments. The 250 samples from the 44 writers was combined and having all the information together, the same amount of data than in the original samples (3498 of test data and 7494 of training data) was extracted but the selection was done randomly. The mean and standard deviation of those 30 experiments are shown in Table 5.12.

Table 5.12: Descriptive results for accuracy for 30 experiments with handwritten dataset.

Copula Model	**Accuracy %** Mean	Std. Dev.
Independent	87.0	0.5
AMH	90.9*	0.5
Clayton	91.1*	0.5
FGM	89.6*	0.6
Frank	93.7*	0.5
Gaussian	94.1*	0.4
Gumbel	93.0*	0.5
Selection	93.5*	0.6

An ANOVA test was performed for these experiments as well and all copulas and copula selection had a statistical difference to independent copula. All copulas and copula selection had a better performance than independent copula, Gaussian copula was the one with the best result with a difference of 7.1%.

6. Discussion

This research had as purpose to incorporate copula functions in the design of a probabilistic classifier and evaluate its performance.

It has been carried out through a series of experiments, an analysis and comparison of the results of a classifier that takes into consideration the dependencies among variables versus a naive Bayes classifier or independent copula which do not take into consideration the dependencies. Next, the main findings of this study will be discussed.

From the results obtained from this research, it can be deduced that the use of copula functions to model dependencies among variables in classification, produces improvements in the performed experiments.

In one of the two datasets used, the use of copulas Clayton, Frank, Gaussian and Gumbel as well as copula selection are reported as cases with a statistical difference compared to independent copula. These results were expected given that the other pair of copulas with which experiments were carried out (AMH and FGM), model weak dependencies, similar to the independent copula.

On the other hand, from the experiments, it can be concluded that not in all cases were obtained better results with the use of the same copulas and that is because each copula models a different type of dependency which suggest that in the analyzed samples there were different types of dependencies. For instance, in the experiments of random sample using 5% of the data as training information, while image "PERSON6" had the highest results with Clayton and Frank copula, "PERSON8" had them with Gaussian and Gumbel copula, the reader is referred to Appendix A to corroborate this information.

The fact that in the mean accuracy of the experiments carried out, there was a higher result for all six copulas, is an indicative that modeling dependencies with the help of copula functions in supervised classification provides an improvement versus not modeling them.

Now, some aspects reviewed of the data obtained during this research will be discussed in detail. There were used two databases to perform the experiments. The first one consisted in 50 images taken from Microsoft repository [26] and the experiments consisted in classifying pixels. From each pixel, three variables (Red, Green and Blue) and two classes (Foreground and Background) were taken into consideration. The second database [15] consisted on handwritten numbers from 0 to 9, in this database there were 16 variables which was coordinate information and 10 classes (0-9).

The database of the images was the most explored in this research since the images provided by the repository were used but experiments were also done with the same images only selecting randomly the pixels for training and test, and a case were the training data was selected by freehand was also simulated.

In the first case, when it was experimented with the images as they are in the repository, the copula with the highest result was the Frank copula with a mean of 87.7% of the pixels correctly classified (accuracy) versus the independent copula with a mean of 79.4% in accuracy.

During the second stage of experiments where the pixels were selected randomly as training and test data, Gaussian copula was the one with highest result in all cases (5, 10, 20, 30, 40 and 50 percent of pixels as training data). For the case of 5% of pixels as training data, the result of Gaussian copula on terms of mean accuracy was 92.4% versus 84.2% of independent copula.

In the case when the training data for foreground and background was selected by freehand, both previously mentioned copulas, Frank and Gaussian, were the ones with the highest results in mean accuracy with 92.2% versus 83.6% of independent copula.

For the results of the 30 experiments performed with the handwritten digits database, the highest result was Gaussian copula with 94% of mean accuracy versus 86.9% in mean accuracy for independent copula.

From this research, it was found that the use of copula functions to model dependencies in supervised classification provides competitive results, worthy of further study.

The general objective and specific objectives were all achieved with satisfactory results. With the results on the experiments, it can be concluded that the research hypothesis was demonstrated for the samples used. Copula functions incorporated to supervised classification improved the results by modeling dependencies with the help of a graphical chain model.

6.1 Limitations of the Study

An unexpected finding was to realize that the mean accuracy on copula selection did not have the highest result. The method to select the best copulas was to find the pair of variables with the highest log-likelihood, this suggested that the result would be higher than using the same copula for all pair of variables in the graphical chain model. It is not clear why it did

not happen and some hypothesis could come up but since it requires more study, the results should be interpreted with caution. However, the results of copula selection were not lower than independent copula, they were always higher with a statistical difference.
The optimization and time of response on the algorithms used was not taken into account. Since this study focused on the evaluation metrics especially on accuracy of the experiments, the optimization and time of response on the algorithms may have not been the best, this is an important aspect for future work especially when comparing the classifier with other methods.
Another limitation was the use of only two databases for the experiments. To have a deeper insight of the advantages of incorporating copula functions in supervised classification it is recommendable to use more databases and perform more experiments.

6.2 Recommendations and Future Work

For future work, it is recommended for the results to be compared with other classification methods, probabilistic or non probabilistic.
Even when the classifiers are compared, it is important to have in mind that every different classifier can be the best option for a different dataset. It would be useful to create a method to identify in which cases the supervised classification with copulas incorporated would be the best option to classify.
It is also recommended to experiment with other databases and to have more variables and classes. One option could be a dataset where the results on independent copula are lower than the ones used for this research given that the samples used in the study already had a high result when not taking into account the dependencies among variables.
As mentioned in the previous section, it would be useful to optimize and improve the response time of the algorithms used.

7. Conclusions

In this study it was proposed to determine the performance of a probabilistic classifier when copula functions are incorporated into supervised classification. The performance improves in comparison with the same probabilistic classifier without taking into consideration the dependencies among variables.

There is a statistical difference for copulas Clayton, Frank, Gaussian, Gumbel and copula selection for the first dataset used in the experiments and there is a statistical difference between all 7 cases (6 copulas and copula selection) versus independent copula for the second one.

This study has demonstrated that modeling dependencies among features provides important information for supervised classification. For the databases used in this work, Gaussian and Frank copula performed very well in most cases.

With the help of a graphical chain model, the most important bivariate dependencies among variables were detected which was an advantage to avoid using all dependencies between each pair of variables and use only the most important ones.

The findings in this research can be helpful for different applications where association among variables can provide important information of the problem.

Bibliography

[1] Ahmad, P., Qamar, S. & Afser, S.. (2015). Techniques of Data Mining In Healthcare: A Review. *International Journal of Computer Applications*, 120 , pp.38-50

[2] Bain, L.J & Engelhardt, M.. (1992). Introduction to Probability and Mathematical Statistics. Belmont, California: Duxbury Press.

[3] Bezak, N., Mikoš, M. & Šraj, M. Water Resources Manage (2014) 28: 2195. https://doi-org.dibpxy.uaa.mx/10.1007/s11269-014-0606-2

[4] Boos, D.D. & Stefanski, L.A. (2013). Likelihood Construction and Estimation. In: Essential Statistical Inference. Springer Texts in Statistics, vol 120. Springer, New York, NY

[5] Braeken, J., Tuerlinckx, F. & De Boeck, P. (2007). Copula Functions for Residual Dependency. Psychometrika 72: 393-411

[6] Brunel, N., Pieczynski, W. & Derrode, S. (2005). Copulas in vectorial hidden markov chains for multicomponent image segmentation. In: ICASSP '05: Proceedings of the 2005 IEEE International Conference on Acoustics, Speech and Signal Processing.717-720

[7] Carrera, D., Santana, R. & Lozano, J. (2016). Vine copula classifiers for the mind reading problem. Progress in Artificial Intelligence

[8] Cook, D., & Swayne, D. F. (2007). Interactive and Dynamic Graphics for Data Analysis. Springer New York,63-100

[9] Dougherty, G. (2013). Pattern Recognition and Classification. Springer New York.

[10] Encyclopedia of Database Systems. Springer US.,pp. 1885-1890 (2009)

[11] Encyclopedia of Machine Learning. Springer US.,p. 716 (2010)

[12] Getting to Know GIMP (2008). In: Beginning GIMP. Apress, pp. 1-18

[13] Härdle W. (1991) *Kernel Density Estimation.* In: Smoothing Techniques. Springer Series in Statistics (pp.43-84). Springer, New York, NY

[14] Kalb I. (2016) Getting Started. In: Learn to Program with Python. Apress, Berkeley, CA

[15] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[16] Liu B., Yang Y., Webb G.I., Boughton J. (2009) A Comparative Study of Bandwidth Choice in Kernel Density Estimation for Naive Bayesian Classification. In: Theeramunkong T., Kijsirikul B., Cercone N., Ho TB. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2009. Lecture Notes in Computer Science, vol 5476. Springer, Berlin, Heidelberg

[17] Mercier, G., Bouchemakh, L. & Smara, Y. (2007). The Use of Multidimensional Copulas to Describe Amplitude Distribution of Polarimetric SAR Data. In: IGARSS 07

[18] Murty, M. N., & Devi, V. S. (2011). Pattern Recognition. Springer London.

[19] Nelsen, R.B. (2006). An Introduction to Copulas. Springer, Heidelberg

[20] O'Farrell, M., Lewis, E., Flanagan, C., Lyons, W. & Jackman, N.. (2005). Comparison of k-NN and neural network methods in the classification of spectral data from an optical fibre-based sensor system used for quality control in the food industry. *Sensors and Actuators B: Chemical*, 111-112, pp. 354-362.

[21] Patton A.J. (2009) Copula–Based Models for Financial Time Series. In: Mikosch T., Kreiß JP., Davis R., Andersen T. (eds) Handbook of Financial Time Series. Springer, Berlin, Heidelberg

[22] Pérez-Díaz, A., Salinas-Gutiérrez, R., Hernández-Quintero, A. & Dalmau- Cedeño, O. (2017). Supervised Classification Based on Copula Functions. In Research in Computing Science(pp. 9-18). Mexico : Instituto Politécnico Nacional

[23] Pfeifer, D. & Nešlehová, J. (2003). Modeling dependence in finance and insurance: the copula approach. Blätter DGVFM. Springer-Verlag 26: 177-191

[24] Peters J.F. (2014) Descriptive Raster Spaces. In: Topology of Digital Images. Intelligent Systems Reference Library, vol 63. Springer, Berlin, Heidelberg

[25] Pigeot I., Klasen S. & Foraita R. (2010) Graphical Chain Models and their Application. In: Kneib T., Tutz G. (eds) Statistical Modelling and Regression Structures. Physica-Verlag HD

[26] Rother, C., Kolmogorov, V., Blake, A. & Brown, M.: Image and video editing. http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm

[27] Sakji-Nsibi, S. & Benazza-Benyahia, A. (2008). Multivariate indexing of multichannel images based on the copula theory. In: IPTA08

[28] Salarpour M., Yusop Z., Yusof F., Shahid S. & Jajarmizadeh M. (2016) Flood Frequency Analysis Based on Gaussian Copula. In: Tahir W., Abu Bakar P., Wahid M., Mohd Nasir S., Lee W. (eds) ISFRAM 2015. Springer, Singapore

[29] Salinas-Gutiérrez, R., Hérnandez-Aguirre, A., Rivera-Meraz, M. J. J. & Villa-Diharce, E. R. (2011). Using gaussian copulas in supervised probabilistic classification. In Soft Computing for Intelligent Control and Mobile Robotics. Springer Berlin Heidelberg,pp. 355-372

[30] Salinas-Gutiérrez R., Hernández-Quintero A., Dalmau-Cedeño O. & Pérez-Díaz Á.P. (2017) Modeling Dependencies in Supervised Classification. In: Carrasco-Ochoa J., Martínez-Trinidad J., Olvera-López J. (eds) Pattern Recognition. MCPR 2017. Lecture Notes in Computer Science (pp. 117-126), vol 10267. Springer, Cham

[31] Ščavnický, M. (2013). A study of applying copulas in data mining. Master's thesis, Charles University in Prague, Prague

[32] Sen, S., Diawara, N. & Iftekharuddin, K. (2015). Statistical Pattern Recognition Using Gaussian Copula. Journal of Statistical Theory and Practice 9(4), 768-777

[33] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231

[34] Slechan, L. & Górecki, J. (2015). On the Accuracy of Copula-Based Bayesian Classifiers: An Experimental Comparison with Neural Networks. In: Núñez, M., Nguyen, N., Camacho, D., Trawiński, B. (eds.) Computational Collective Intelligence: 7th International

Conference, ICCCI 2015, Madrid, Spain, September 21-23, 2015, Proceedings, Part I. Lecture Notes in Computer Science, vol. 9329, pp. 485-493. Springer International Publishing, Cham

[35] Stitou, Y., Lasmar, N. & Berthoumieu, Y. (2009). Copulas based multivariate gamma modeling for texture classification. In: ICASSP '09: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing.1045-1048. IEEE Computer Society, Washington, DC, USA

[36] The Concise Encyclopedia of Statistics. New York: Springer New York.(2008)

[37] Tyagi V. (2017) Similarity Measures and Performance Evaluation. In: Content-Based Image Retrieval. Springer, Singapore

[38] Weiß, G. (2011). Copula parameter estimation by maximum-likelihood and minimum distance estimators: a simulation study. Computational Statistics 26(1), 31-54

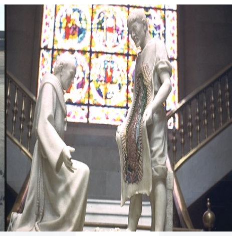
[39] Williams G. (2011). Support Vector Machines. In: Data Mining with Rattle and R. Use R. Springer, New York, NY, 293-304

[40] Zheng J. & Tang Y. (2005) One Generalization of the Naive Bayes to Fuzzy Sets and the Design of the Fuzzy Naive Bayes Classifier. In: Mira J., Álvarez J.R. (eds) Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach. IWINAC 2005. Lecture Notes in Computer Science, vol 3562. Springer, Berlin, Heidelberg

A. Accuracy per Image Tables

Images


21077


24077


37073


65019


69020


86016

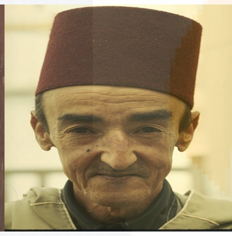

106024


124084


153077


153093


181079


189080


208001


209070


227092


271008


304074


326038


376043


388016


BANANA1


BANANA2


BANANA3


BOOK


BOOL

BUSH CERAMIC CROSS DOLL ELEPHANT

FLOWER FULLMOON GRAVE LLAMA MEMORIAL

MUSIC PERSON1 PERSON2 PERSON3 PERSON4

PERSON5 PERSON6 PERSON7 PERSON8 SCISSORS

SHEEP STONE1 STONE2 TEDDY TENNIS

Independent

Accuracy for each image with independent copula, the results are presented in percentages. RS stands for Random Sample.

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
21077	82.2	82.5	79.2	84.4
24077	62.6	62.3	65.7	65
37073	82.8	82.2	80.7	85.6
65019	78.4	79.4	77.5	71.2
69020	73.5	73.3	72.6	72.6
86016	93.8	93.7	93.7	89.9
106024	62.3	63.2	56.6	57.3
124080	96.8	96.1	95.5	92
153077	84.9	85.3	86.5	72.2
153093	66.4	66.3	68.2	68.8
181079	81.9	81.8	82.4	72.8
189080	89.9	90.2	89	88.7
208001	90.7	90.7	88.6	89.9
209070	72.1	74	75.8	69.8
227092	87.6	86.9	86.6	57
271008	68.5	69.9	75.2	83.9
304074	71.3	71.3	58.4	62.9
326038	81.7	80.4	82.9	73.9
376043	76	77.1	78.2	67.7
388016	89.8	89.5	79	79.8
BANANA1	93.1	93.1	92.2	76.1
BANANA2	87.1	87	84.1	76.7
BANANA3	91	90.8	91	82
BOOK	92.6	92.4	87.9	77
BOOL	87.1	86.9	91.7	88.5

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
BUSH	67.5	69.4	59.1	60.6
CERAMIC	94.5	94.7	92.3	88.5
CROSS	99	99	99.1	97
DOLL	83.5	83.8	82.1	77.7
ELEPHANT	94	93.9	94.7	86
FLOWER	90.1	89.7	94.6	94.2
GRAVE	77.9	77.5	80.4	93.5
LLAMA	82.7	83.2	82.9	75.4
MEMORIAL	82.9	82.9	85.3	72.2
MOON	99.9	100	99.8	92.1
MUSIC	92.3	92.1	90.9	90.2
PERSON1	82.4	82.5	81.5	75.7
PERSON2	83.1	83.8	85.7	75.6
PERSON3	87.3	87.4	90.7	81.8
PERSON4	69.3	69.1	61.8	72.3
PERSON5	90.5	90.1	90.9	83.1
PERSON6	81.9	80.4	79.4	75
PERSON7	78.4	79.2	80.4	75.5
PERSON8	59.8	59.4	59.9	58.1
SCISSORS	96.5	96.4	95.9	89.4
SHEEP	92	92.1	91.1	87.3
STONE1	95.6	95.7	96.6	90.4
STONE2	98.3	98.4	98.4	93.6
TEDDY	99.1	99.2	99	97.6
TENNIS	87	86.8	88.5	82.7

AMH

Accuracy for each image with AMH copula, the results are presented in percentages. RS stands for Random Sample.

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
21077	84.3	86.4	81.9	80.9
24077	53.1	55.3	59.1	68.7
37073	88.3	87.7	86.8	85.9
65019	80.6	82.1	80.5	75.6
69020	76	76.1	77.4	87.9
86016	95.1	95.2	95.2	92
106024	72	72.4	60.2	56.1
124080	96.7	96.1	95.6	92.1
153077	87.7	87.6	87.6	74.9
153093	71.7	72	71.9	81.6
181079	84	84.5	84.3	76.5
189080	89.2	89.6	88.6	87.3
208001	90.8	90.7	90.1	91.4
209070	80.2	79.6	84.1	72.5
227092	89.5	88.7	89.1	81.2
271008	67.5	68.7	73.9	84.5
304074	69.9	68.7	70.3	61.9
326038	87	86	82.9	75.2
376043	80.9	82.2	82.1	71.5
388016	93.4	94.1	95.7	78.8
BANANA1	96	96	94.8	81.3
BANANA2	88.6	88.6	86.1	77.9
BANANA3	94.8	94.7	94.9	86.3
BOOK	97.7	97.7	95.5	83.2
BOOL	86.5	82.8	87.4	89.5

NAME	**RS 5%**	**RS 50%**	**FREEHAND**	**DATABASE**
BUSH	84	84.3	77.5	71.2
CERAMIC	96.6	96.7	95.1	93.1
CROSS	99	99	99.1	97.3
DOLL	90	89.4	87.2	86.5
ELEPHANT	94.1	94.1	94.8	86.4
FLOWER	93.8	93.4	96	96.8
GRAVE	77.3	76.9	79.6	95.7
LLAMA	87.1	87.4	89.6	76.6
MEMORIAL	84.9	85	87.4	73.8
MOON	99.9	100	99.5	92.1
MUSIC	93.2	93.2	90.7	91.1
PERSON1	86.7	86.3	85.6	74
PERSON2	86.8	89	89.6	78.9
PERSON3	92.7	92.9	96.7	86.6
PERSON4	73.5	71.4	66	75.3
PERSON5	93.1	93	94.2	89.4
PERSON6	80.1	78.2	78	78.6
PERSON7	84.5	84.5	84.3	83.4
PERSON8	59.3	59.1	58.7	69.4
SCISSORS	96.5	96.3	95.4	89.4
SHEEP	95.8	95.8	96.1	92.3
STONE1	96.4	96.5	97	94.5
STONE2	98.5	98.6	98.6	97.4
TEDDY	99	99.2	99.1	97.6
TENNIS	86.3	88.7	89	82.5

Clayton

Accuracy for each image with Clayton copula, the results are presented in percentages. RS stands for Random Sample.

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
21077	83.5	84.9	79.5	80.5
24077	75.7	70.8	75.3	74.4
37073	90.8	90.8	89.1	86.4
65019	89.9	89.3	87.9	75.3
69020	87.2	87	86.3	91.4
86016	96.9	96.9	96.2	90.3
106024	84.6	86.9	89.3	82.3
124080	97.7	97.7	97.1	93
153077	92	92	92.4	82.3
153093	87.3	87.6	91.2	87.5
181079	91.4	91.3	91.7	84.1
189080	95.1	95.1	94.3	94.2
208001	90.5	90.3	91.9	90.2
209070	73.5	75.1	77.6	75.5
227092	95.7	95	96.5	90.2
271008	82.6	82.3	85	69
304074	79.1	80.1	75.1	63
326038	86.4	86.1	82.9	79.9
376043	90.9	91.7	90.6	80.2
388016	97.4	97.3	96.2	97.6
BANANA1	96.1	95.7	92.7	80
BANANA2	92.8	92.9	91.1	81.1
BANANA3	96.6	96.7	96.8	85.4
BOOK	98.8	98.7	95.7	82.7
BOOL	92	90.3	93.8	91.5

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
BUSH	86.9	87	85.5	73.7
CERAMIC	97.4	97.3	95.8	93.7
CROSS	92.6	93.3	99	97.2
DOLL	95.9	95.9	97.3	95
ELEPHANT	96.6	96.4	97.8	93.1
FLOWER	95.7	95.6	96.5	97.3
GRAVE	79.5	78.8	78.1	74.6
LLAMA	88.1	89.1	90.3	84.8
MEMORIAL	80.2	79	87.1	76.1
MOON	99.9	100	99.7	92.1
MUSIC	94.2	94.3	91.7	92.5
PERSON1	96.2	95.7	95.2	87.6
PERSON2	96.5	96.9	96.8	95.3
PERSON3	84	92.1	94.8	87.8
PERSON4	86.3	86.9	89.6	80.5
PERSON5	88.2	87.5	92.4	90.7
PERSON6	91.3	90.6	90.5	78.5
PERSON7	89.7	89.2	90.3	92.9
PERSON8	89.2	89.1	88.7	75.8
SCISSORS	95	96.2	91.9	93.5
SHEEP	98	98.1	97.1	97.4
STONE1	99.7	99.7	99.1	98.8
STONE2	99.3	99.2	99.3	95.7
TEDDY	86.5	99.4	99.1	97.2
TENNIS	91.2	90.5	91.2	73.1

FGM

Accuracy for each image with FGM copula, the results are presented in percentages. RS stands for Random Sample.

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
21077	83.8	87.1	83.1	82.1
24077	60.1	59.9	62.6	68.4
37073	83.2	82.6	82.9	86.1
65019	79.7	80.8	78.2	76.4
69020	75	75	74.9	75.2
86016	93.7	93.7	93.6	90
106024	59.6	60.9	52.1	56.4
124080	96.9	96.1	95.5	92.7
153077	87	87.3	87.1	73
153093	66.8	67	69.3	71.3
181079	82.7	82.6	82.3	73.1
189080	90	90.4	89	89.1
208001	91.4	91.3	90	90
209070	76.6	77.7	79.5	71.3
227092	87.7	87.1	88	74
271008	66.5	66.9	72.1	84.5
304074	72	71.6	59.7	62.3
326038	84.3	83.4	86.4	74.6
376043	77.8	79.1	80	67.9
388016	91.8	92.5	90.6	81.7
BANANA1	95	94.9	93.7	78.2
BANANA2	88.8	88.8	86.3	78.4
BANANA3	92.8	92.8	91.7	81.2
BOOK	95.1	94.9	90.3	78.8
BOOL	87.1	86.7	92.4	88.4

NAME	**RS 5%**	**RS 50%**	**FREEHAND**	**DATABASE**
BUSH	74.4	76.1	63.9	65.3
CERAMIC	95	95.2	92.8	89.3
CROSS	99	99	99.1	97
DOLL	83.6	83.9	83.4	80.3
ELEPHANT	94	93.9	94.6	85.9
FLOWER	92.9	92.6	96	96.7
GRAVE	77.6	77.2	79.9	94.2
LLAMA	85.5	85.9	86	75.5
MEMORIAL	84.8	84.9	87.2	71.7
MOON	99.9	100	99.8	92.1
MUSIC	93	92.9	91.3	91.3
PERSON1	85.2	85.8	85.8	75.9
PERSON2	86	86.9	88	75.6
PERSON3	89.1	89.1	92.9	82.5
PERSON4	71.2	70.4	65.1	73
PERSON5	90.4	90.1	91.8	87
PERSON6	81.5	80.1	80.8	76.8
PERSON7	82.6	82.8	84.1	80.5
PERSON8	59.5	59	58.4	65.7
SCISSORS	96.4	96.4	95.9	89.4
SHEEP	94.1	94.2	93.7	88.9
STONE1	96.5	96.5	96.9	90.8
STONE2	98.4	98.4	98.4	94.8
TEDDY	99.1	99.2	99	97.6
TENNIS	87.2	90	89.9	82.7

Frank

Accuracy for each image with Frank copula, the results are presented in percentages. RS stands for Random Sample.

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
21077	89	88	90.8	86.4
24077	76.7	76.1	78.5	76.4
37073	90.1	90.2	88	88.3
65019	88.9	89.2	87.9	77.3
69020	87.7	87.7	87.7	87.6
86016	95.8	95.9	96.5	92.1
106024	83.1	83.2	86.7	80.3
124080	98.4	98.5	97.2	92.3
153077	91.7	91.5	92	78.6
153093	93.3	93.3	90.6	88.6
181079	93	93.2	93.2	87.6
189080	94.9	94.6	93.6	89.6
208001	91	91.3	91.9	90.7
209070	74.9	75.2	74.3	73.7
227092	98.4	98.4	96.2	89.6
271008	71	70.9	80.4	86.4
304074	77.3	78.7	81.2	71.9
326038	79	80	81.7	77.4
376043	90.8	92.1	91.4	83.3
388016	96.9	96.4	95.8	94.1
BANANA1	96.1	96.1	94.9	80.5
BANANA2	95.5	95.4	94.2	91
BANANA3	97.4	97.3	97.2	88.2
BOOK	97.8	97.8	96.8	83.4
BOOL	92.6	92.4	94.3	89.6

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
BUSH	85.7	86.5	86.3	76.4
CERAMIC	95.6	95.7	93.5	90.2
CROSS	98.9	98.9	98.8	96.8
DOLL	96.9	96.9	97.6	94.2
ELEPHANT	94	93.8	95.3	86.1
FLOWER	95.9	95.8	96.3	97.1
GRAVE	79.1	78.4	80.4	93.6
LLAMA	91.5	91.9	95	86
MEMORIAL	81.6	80.9	85.6	77.7
MOON	99.9	100	98.5	92.1
MUSIC	93.6	93.4	93.3	93.5
PERSON1	97.1	97.4	97.2	94.9
PERSON2	98.3	98.2	98.4	97.1
PERSON3	92.4	92.5	96.3	94.3
PERSON4	91	91	93.1	82.3
PERSON5	90.8	90.7	95.8	92.1
PERSON6	91	90.9	88.8	82.8
PERSON7	92.6	92.4	94.5	94.8
PERSON8	89.6	88.8	90.9	80.8
SCISSORS	94.1	94.5	93.7	94
SHEEP	98.4	98.5	98.6	97
STONE1	99	99.1	98.5	91.5
STONE2	98.7	98.8	98.9	97.9
TEDDY	99.2	99.3	99	97.5
TENNIS	94.6	94.4	93.4	78.2

Gaussian

Accuracy for each image with Gaussian copula, the results are presented in percentages. RS stands for Random Sample.

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
21077	88.7	88.4	90	82.6
24077	72.8	73.2	70.7	76.6
37073	92.6	92.6	89.6	88.6
65019	86.1	86.6	85.4	79.9
69020	88.6	88.6	87.8	87.3
86016	97.8	97.9	98	92.2
106024	92.4	86.5	81.3	77.9
124080	98.9	98.9	98	93.7
153077	92.6	92.3	92.7	82.6
153093	90.1	89.9	90.2	88.5
181079	93.6	93.6	93.7	87.8
189080	95.4	95.2	95.4	56.1
208001	91.3	91.4	91.7	90.1
209070	82.5	82.6	83.1	81.5
227092	98.2	98	97.8	87.9
271008	68.6	67.9	72.1	87.8
304074	78.3	79	74.9	69.2
326038	87.9	87.9	88	81.8
376043	94.1	94.1	93.9	82.9
388016	97.4	97.5	97.5	96.2
BANANA1	96.1	96	94.6	81.2
BANANA2	95.1	95.1	93.5	90.7
BANANA3	97.5	97.5	97.3	81.7
BOOK	99.2	99.2	98.3	84.5
BOOL	92.7	92.2	93.7	90.2

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
BUSH	86.9	87.5	87.7	76.1
CERAMIC	96.1	96.2	94.3	92.9
CROSS	98.9	98.8	98.8	44.8
DOLL	97	97	97.6	96.4
ELEPHANT	95.9	95.7	96.8	64.4
FLOWER	95.6	95.5	96.2	97
GRAVE	80.6	79.9	81	94
LLAMA	91.7	92.4	94.9	86.1
MEMORIAL	81.7	81.1	82.5	79.3
MOON	99.9	100	99.6	87.7
MUSIC	93.1	93	95.2	93.8
PERSON1	97.8	98	97.7	94.6
PERSON2	98.4	98.2	97.8	96.9
PERSON3	94.1	93.9	97.3	93.6
PERSON4	90.8	90.3	93	82.6
PERSON5	90.9	91	93	90.2
PERSON6	89.5	88.1	87.8	82.2
PERSON7	93.9	94.6	95.7	94.8
PERSON8	91.5	90.8	90.6	83.3
SCISSORS	95.7	96.6	92.8	94.4
SHEEP	98.2	98.2	98.4	96.6
STONE1	99.5	99.5	98.5	98.2
STONE2	99.3	99.3	99.4	98.3
TEDDY	99.3	99.3	99	97.3
TENNIS	94.6	93.9	93.7	83.7

Gumbel

Accuracy for each image with Gumbel copula, the results are presented in percentages. RS stands for Random Sample.

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
21077	89.5	89.2	91.5	84.4
24077	75.1	74.8	73.3	78.1
37073	91.5	91.3	88.7	87.6
65019	87.9	88	87.2	79.9
69020	88.4	88.3	88.6	81.7
86016	97.5	97.6	97.8	93.6
106024	88.8	85.6	87.1	74.8
124080	99	99	98.1	93.1
153077	92	91.7	92.5	80
153093	91.2	90.2	87.6	87.4
181079	93.4	93.5	93.9	87.8
189080	95.2	95	93.7	56
208001	90.4	90.7	91.7	88.8
209070	81.9	82	83.9	80
227092	98.1	97.8	95.8	86.6
271008	70.1	69.8	72.4	87.3
304074	74.7	76	75.2	69.7
326038	85.2	85.6	85.7	80.5
376043	93.1	93.1	94.5	83
388016	97	96.7	96.6	94.9
BANANA1	95.7	95.7	94.9	80.4
BANANA2	95.6	95.4	94.3	92.1
BANANA3	97.6	97.5	97.4	88
BOOK	98.8	98.8	98.2	84.1
BOOL	92.6	92.3	94.3	89.8

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
BUSH	85.3	86	88.1	78
CERAMIC	95.7	95.8	93.8	90.7
CROSS	98.9	98.9	99	79.1
DOLL	97.1	97.1	97.4	95.5
ELEPHANT	95.9	95.9	96.5	89.3
FLOWER	95.9	95.9	96.2	97
GRAVE	82.7	82.3	83.5	93.7
LLAMA	92.2	92.6	81.9	82.4
MEMORIAL	81.4	80.9	81.6	77.4
MOON	99.8	100	98.3	87.7
MUSIC	94	93.9	94.9	94.3
PERSON1	96.1	96.6	96.4	90.9
PERSON2	98	97.7	98.1	96.2
PERSON3	93.8	93.6	87.4	94.7
PERSON4	91.5	91.3	93.5	86.8
PERSON5	93.2	93.3	95.3	91.1
PERSON6	89	88	88.3	77.5
PERSON7	93.4	93.9	95.6	93.5
PERSON8	91	90.2	90.1	81.6
SCISSORS	93.1	93.1	92.4	94.5
SHEEP	98.2	98.2	98.4	94.9
STONE1	99.2	99.1	97.7	94.8
STONE2	98.8	98.9	99.1	98.7
TEDDY	99.3	99.3	99	97.6
TENNIS	95	94.8	94.2	89.9

Copula Selection

Accuracy for each image with copula selection, the results are presented in percentages. RS stands for Random Sample.

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
21077	88.1	88.9	91.3	86
24077	70.3	71.9	75	77.4
37073	91.5	91.3	88.2	88.1
65019	89.3	89.7	90	79.7
69020	88.8	88.7	88.6	85.4
86016	97.8	97.9	98	92.2
106024	93	90.1	87.4	79.9
124080	99	99	98.1	93.1
153077	91.9	91.7	93	80.1
153093	91.5	90.5	90.5	88.3
181079	93.6	93.6	94.2	88.4
189080	96.6	96.6	94.7	56
208001	90.8	90.9	91.4	86.8
209070	78.3	79.1	75.2	73.3
227092	87.7	88	97.6	91.2
271008	70.1	72.2	77.2	88.6
304074	79.1	79.6	74.9	70.6
326038	87.8	86.3	87.5	80.9
376043	94.3	94.3	95.2	82.1
388016	97.5	97.2	97	95.9
BANANA1	96.7	96.6	95.1	81.5
BANANA2	95.7	95.6	94.6	91
BANANA3	97.6	97.7	97.6	88.2
BOOK	99	99.1	98.1	84.1
BOOL	92.6	91.5	93.2	91.3

NAME	RS 5%	RS 50%	FREEHAND	DATABASE
BUSH	85.3	86	88.1	78.3
CERAMIC	97	97	94.5	93.7
CROSS	92.3	93.2	98.8	44.8
DOLL	97.2	97.2	97.5	94.3
ELEPHANT	97.6	97.7	97.9	67.9
FLOWER	96	96	95.9	96.8
GRAVE	83	82.6	83.5	93.9
LLAMA	92.5	91.7	94.5	85.4
MEMORIAL	84.5	84	86.6	78.9
MOON	99.9	100	98.5	92.1
MUSIC	93.9	93.7	94.8	94.3
PERSON1	97.7	97.7	97.6	94.6
PERSON2	98.6	98.5	98.6	97.9
PERSON3	87	92.2	96.8	95.4
PERSON4	89.9	89.7	91.9	85.9
PERSON5	91.1	91.1	95.3	92.8
PERSON6	91.2	90.9	89.8	82.8
PERSON7	94	94.3	95.1	94.2
PERSON8	90.5	88.3	90.7	80.8
SCISSORS	94.5	94.9	93.5	94.6
SHEEP	98.3	98.5	98.6	96.5
STONE1	99.6	99.6	98	98.7
STONE2	99.3	99	99.4	98.3
TEDDY	99.2	99.3	98.7	96.5
TENNIS	94.7	94.4	93.3	78.3