

# Clasificación de malformaciones craneales causadas por craneosinostosis primaria utilizando kernels no lineales

Salvador Ruiz-Correa\*  
Yerania Campos-Silvestre\*

\* Departamento de Cómputo Matemático.  
Centro de Investigaciones Matemáticas.  
Universidad de Guanajuato.

Correspondencia:

Dr. Salvador Ruiz Correa  
Jalisco s/n, Col. Valenciana, Guanajuato,  
Guanajuato, México, D.F. C.P. 36240, E-mail:  
src@cimat.mx

Artículo recibido: 16/octubre/2009

Artículo aceptado: 16/marzo/2010

## RESUMEN

La *craneosinostosis primaria* (CP) es una condición patológica asociada a la fusión prematura de una sutura del calvario. La fusión temprana produce malformaciones craneales severas que se asocian frecuentemente con discapacidades cognitivas, del lenguaje, el aprendizaje y el comportamiento social en general. Por estas razones, la CP representa un área de investigación importante que requiere de métodos efectivos para caracterizar la morfología del complejo craneofacial. En este trabajo, se estudia el desempeño de un algoritmo que combina en forma novedosa *kernels no lineales*, *conurrencias de rasgos craneales de forma*, un método de selección de variables y un método estándar de reducción de dimensionalidad para clasificar malformaciones causadas por CP en una población de infantes afectados, utilizando imágenes de tomografía computarizada (TC). Las imágenes fueron obtenidas de 102 estudios de cráneos con sinostosis sagital, 42 cráneos con sinostosis metópica, 12 cráneos con sinostosis unicoronal y 65 cráneos de individuos sanos (población control). El estudio cuantitativo de los resultados obtenidos sugiere que el método propuesto es efectivo, ya que alcanza tasas de clasificación superiores al 94%.

**Palabras clave:** Craneosinostosis primaria, craneosinostosis metópica, craneosinostosis sagital, malformación craneal, análisis de forma, descriptores de forma, complejo craneofacial, clasificación.

## ABSTRACT

Single-suture craniosynostosis (SSC) is the pathologic condition of premature fusion of a calvarial suture. Premature fusion produces significant cranial deformities and is associated with an increased risk of cognitive deficits and neurobehavioral impairments. For these reasons, SSC represents an important area of research that requires effective methods for characterizing cranial morphology. In this paper we evaluate a new approach that combines the use of nonlinear kernels, co-occurrences of skull shape features, a new feature selection process and standard nonlinear dimensionality reduction techniques, as a means to classify cranial malformations due to SSC using computed tomography (CT) imaging. CT images were obtained from CT studies of 102 sagittal synostosis crania, 42 metopic synostosis crania, 12 unicoronal synostosis crania and 65 nonsynostotic skulls. We validate our approach with an extensive series of experiments and show that our proposed approach outperforms the classification performance of previously published techniques, achieving classification rates above 95%.

**Key words:** Craniosynostosis, cranial malformations, sagittal synostosis, metopic synostosis, cranial malformation, shape analysis, craniofacial complex, classification.

## INTRODUCCIÓN

La craneosinostosis se define como la fusión temprana de una o más suturas del calvario (sagital, metópica, coronal o lambdoide) que normalmente separan las placas de hueso que forman la bóveda del cráneo<sup>1</sup>. En infantes que se desarrollan con normalidad, el cerebro crece gradualmente dentro del cráneo gracias a que las suturas se encuentran abiertas. En niños que padecen craneosinostosis, el cierre precoz de una o más suturas restringe la expansión del cráneo en la dirección perpendicular a la(s) sutura(s) cerradas y produce un crecimiento compensatorio en la dirección de las suturas que se encuentran abiertas. Este proceso anormal altera severamente la forma del cráneo e incrementa de manera significativa la probabilidad de deficiencias cognitivas, así como las discapacidades del lenguaje y el aprendizaje en infantes afectados. Actualmente, el único tratamiento efectivo para paliar los efectos de la craneosinostosis consiste en remodelar quirúrgicamente el cráneo durante las primeras semanas de vida.

La CP es la forma más común de craneosinostosis. Se presenta con una prevalencia de aproximadamente 1 en 2,500 nacimientos<sup>2</sup>, y está asociada con un incremento significativo en el riesgo de discapacidades cognitivas, del lenguaje, el aprendizaje y el comportamiento social<sup>3</sup>.

La craneosinostosis de las suturas sagital, unilateral coronal y metópica ocurren con mayor frecuencia. La *craneosinostosis sagital* se manifiesta durante el nacimiento como un cráneo alargado y angosto (*escafocefalia*) deformado, con protuberancias bifrontales y occipitales. La *craneosinostosis unilateral*, se caracteriza por un cráneo asimétrico y sesgado (*plagiocefalia*) que presenta además un aplastamiento característico de la frente ipsilateral a la sutura fusionada y una protuberancia compensatoria frontal contralateral localizada al lado opuesto. La *craneosinostosis metópica* produce una cabeza triangular (*trigonocefalia*), caracterizada por una cresta que sobresale de la frente, un angostamiento frontotemporal y un diámetro biparietal anormal<sup>4</sup>.

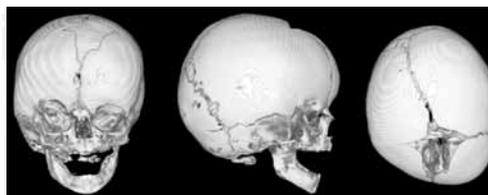
La CP con sus características anatómicas y neurofisiológicas, representa un área importante de investigación e intervención clínica para es-

pecialistas craneofaciales, radiólogos y psicólogos pediátricos. La craneosinostosis primaria es un defecto del nacimiento relativamente común que se diagnostica con frecuencia en instituciones hospitalarias y centros de atención neonatal. La incidencia de craneosinostosis en México no ha sido caracterizada con precisión, sin embargo, es común encontrar casos de craneosinostosis severa en la práctica clínica.

Desafortunadamente, la escasez de métodos cuantitativos para caracterizar la morfología del cráneo ha frenado significativamente el avance en el estudio de las causas de la craneosinostosis y su relación con la forma del cráneo (fenotipo), genes patogénicos candidatos y la exposición a contaminantes mutágenos<sup>5</sup>.

La escasez de métodos para cuantificar la anatomía de un paciente afectado no se debe a la falta de metodologías de imagenología, que hoy en día permiten adquirir información anatómica de manera precisa. La estructura ósea de las suturas de un cráneo sinostótico anormal se pueden apreciar con gran detalle por medio de técnicas de TC que utilizan sistemas compuestos por múltiples detectores y técnicas avanzadas de reconstrucción volumétrica y de despliegue (Figura 1).

En la práctica clínica, la evaluación imagenológica de la craneosinostosis es esencialmente descriptiva y cualitativa. La interpretación médica formal del fenotipo de un paciente se limita usualmente a la descripción subjetiva de la forma dominante del cráneo expresada en términos tales como escafocefalia, trigonocefalia o plagiocefalia. Las alteraciones a la anatomía craneal pueden describirse de manera cualitativa, pero no pueden compararse cuantitativamente con formas del cráneo que son similares o diferentes, con el objeto



**Figura 1.** Reconstrucción volumétrica del cráneo de un paciente afectado con craneosinostosis unicoronal. La reconstrucción se realizó a partir de datos de TC por medio de una estación de trabajo especializada.

de evaluar diferencias entre una clase o clases de cráneos distintos. En el ámbito de la investigación, la mayoría de los estudios cuantitativos de la forma del cráneo, generados a partir de métodos de imagenología médica, se han limitado al registro de mediciones simples tales como el diámetro de la foramina o la distancia entre dos suturas específicas, dejando de un lado la vasta mayoría de la información anatómica disponible<sup>6,7</sup>. Metodologías más elaboradas para cuantificar el complejo craneofacial, tales como la morfometría geométrica<sup>8</sup>, solamente comparan mediciones entre un número limitado de referencias anatómicas que ignoran rasgos que pueden ser importantes para evaluar integralmente defectos anatómicos de importancia. Por esta razón, el desarrollo de métodos cuantitativos para clasificar malformaciones causadas por craneosinostosis primaria es un paso importante en el desarrollo de técnicas cuantitativas para evaluar la morfología craneal<sup>9</sup>.

En este trabajo se presenta un estudio en el que se mide el desempeño de un algoritmo novedoso que combina kernels no lineales, concurrencias de rasgos craneales de forma, un método de selección de variables y un algoritmo estándar de reducción de dimensionalidad para caracterizar y clasificar malformaciones causadas por CP. Este trabajo complementa investigaciones previas en las que se utilizan funciones de clasificación basadas en un Kernel lineal. La técnica de clasificación que se presenta en este trabajo propone que una clase particular de formas craneales puede representarse por medio de patrones característicos de rasgos geométricos denominados *rasgos craneales de forma (RCF)*. La suposición principal es que los *patrones de concurrencia* de los *RCF* se preservan, estadísticamente, en las diferentes instancias de una clase de formas. Se asume también que clases diferentes pueden compartir *RCF*, sin embargo, los patrones de concurrencia de las clases deben de ser diferentes entre sí para que sea posible efectuar la clasificación con tasas de error razonables. Se sugiere que las diferencias en los patrones de concurrencia de rasgos refleja la forma en que propiedades locales de una clase de cráneos se organizan para definir algunas de sus características geométricas globales.

La metodología de clasificación que se presenta incorpora un algoritmo de selección de variables que está basado en una prueba estadística de permutación. El algoritmo es capaz de descubrir patrones distintivos de *RCF* que caracterizan a cada una de las formas de cráneos malformados. El

proceso de selección de variables es seguido por un algoritmo estándar de reducción de dimensionalidad que permite el cálculo de descriptores de forma (DF). Estos descriptores permiten construir funciones de clasificación que generalizan correctamente, es decir, que son capaces de predecir correctamente la clase de forma de un cráneo nunca antes visto.

Las funciones de clasificación de este trabajo se construyeron por medio de *máquinas de soporte vectorial (SVMs)*, por sus siglas en inglés) que utilizan un Kernel no lineal. Con el objeto de validar el funcionamiento de la metodología de clasificación que se propone, se realizaron una serie de experimentos que permiten comparar el desempeño de clasificadores basados en kernels no lineales, particularmente aquéllos derivados de métricas Hilbertianas, con respecto al desempeño de clasificadores que utilizan un Kernel lineal. La razón que motiva este estudio comparativo consiste en disminuir las tasas de error de los clasificadores lineales. Es importante mencionar que, en general, los clasificadores lineales tienen una probabilidad menor de sobreajustar los datos que las diferentes familias de clasificadores no lineales. Sin embargo, en la práctica se ha observado que el uso combinado de SVMs, kernels no lineales y metodologías de validación cruzada pueden disminuir tasas de error de clasificación y las tasas marginales de clasificación de manera significativa, sin sobreajustar los datos.

Este trabajo está organizado de la siguiente manera: La sección antecedentes describe algunas de las investigaciones previas relacionadas con la clasificación de formas anatómicas, en particular, aquellas que utilizan descriptores numéricos de forma. Esta sección también presenta el método de reducción de dimensionalidad que se utilizó en el estudio comparativo. La sección propuesta metodológica desarrolla la metodología de clasificación de formas craneales afectadas por craneosinostosis. Los resultados del estudio experimental se describen en la sección resultados. El artículo concluye con una discusión de las lecciones aprendidas y direcciones futuras de investigación derivadas de este trabajo. Una descripción detallada de los resultados obtenidos con Kernel lineales que motivaron el desarrollo del presente trabajo se describen en<sup>10</sup>.

## ANTECEDENTES

En años recientes se han publicado un número significativo de investigaciones enfocadas al desarrollo de tecnologías para clasificar formas. Estas

tecnologías se han utilizado para resolver eficazmente problemas en las áreas de imagenología médica y visión computacional. El contexto de forma<sup>11</sup>, modelos de estados latentes<sup>12</sup>, gráficas de choque<sup>13</sup>, FORMS<sup>14</sup> y modelos activos de forma<sup>15</sup>, son algunas de las metodologías más exitosas que se han creado para caracterizar y clasificar formas. Desafortunadamente, estos métodos utilizan patrones predefinidos de forma y/o representaciones de alta dimensionalidad que no son apropiados para aplicaciones en investigaciones en craneosinostosis.

Métodos que utilizan las deformaciones de un pequeño grupo de patrones para representar la forma de un objeto no pueden aplicarse directamente porque formas craneales prototípicas en craneosinostosis no han sido definidas cuantitativamente. Si bien es cierto que las variaciones morfológicas en una población pueden dar origen a la existencia de subclases de forma (tal y como lo sugieren Captier y sus colaboradores<sup>16</sup>), no hay evidencia biológica suficiente que sustente la existencia de formas craneales craneosinostóticas patrón que puedan utilizarse para definir subfenotipos.

Las representaciones de forma de baja dimensionalidad son muy útiles en investigaciones en craneosinostosis, porque el tamaño de la población en estudios típicos es relativamente pequeña (del orden de  $\sim 10^2$  muestras). Las representaciones compactas de forma incrementan la probabilidad de que modelos de clasificación y regresión construidos para caracterizar formas craneales generalicen apropiadamente<sup>17</sup>. La creación de modelos efectivos de clasificación y regresión es fundamental en el estudio de las asociaciones que pueden existir entre la morfología craneal y variables biológicas de interés en la medicina craneofacial<sup>9</sup>.

### Descriptores numéricos de forma

Las técnicas tradicionales para cuantificar formas anatómicas incluyen la morfometría, codificación en cadena, factores de forma, momentos invariantes, concavidad fraccional, wavelets de Gabor y descriptores de Fourier<sup>18</sup>. Es bien sabido que métodos basados en factores de forma y momentos invariantes son medidas muy convenientes debido a que su dimensionalidad es baja. Sin embargo, su poder descriptivo es limitado. Por esta razón, es mucho más frecuente encontrar investigaciones que utilizan métodos basados en descriptores de Fourier y wavelets de Gabor<sup>18</sup>. Recientemente se han de-

sarrollado métodos elaborados más efectivos para clasificar formas complejas a partir de contornos. El contexto de forma y los modelos latentes de forma, por ejemplo, se han aplicado con mucho éxito en problemas de visión computacional y sistemas de recuperación de imágenes<sup>11</sup>.

Los descriptores de Fourier y métodos basados en morfometría han sido utilizados para clasificar cráneos afectados por craneosinostosis (cráneos sinostóticos). Por ejemplo, Richtsmeier y sus colegas han combinado análisis de matrices Euclidianas de forma, métodos de máxima verosimilitud y funciones de disimilaridad para clasificar malformaciones craneales a partir de puntos de referencia canónicos del cráneo<sup>9</sup>. Los resultados obtenidos con esta metodología producen tasas de error de clasificación del 36 y 18% para cráneos sagitales y metópicos, respectivamente. Esta metodología motivó el desarrollo del espectro craneal<sup>9</sup>, y más recientemente, nuestros descriptores simbólicos de forma, que permiten la discriminación de formas sinostóticas con tasas de error del 10%<sup>20</sup>. La combinación de la tecnología de descriptores simbólicos de forma, una técnica de selección de variables y un algoritmo estándar de reducción de dimensionalidad permitieron el desarrollo de una metodología basada en concurrencias de rasgos craneales que es capaz de reducir la tasa de error de clasificación significativamente<sup>10</sup>.

### Métodos de reducción de dimensionalidad

La mayoría de los métodos de análisis de patrones requieren de una técnica de reducción de dimensionalidad para lograr resultados estadísticamente significativos. Los algoritmos de reducción de dimensionalidad tienen como objetivo explorar y descubrir representaciones compactas de datos de dimensionalidad alta. Existen una variedad de métodos para extraer información relevante por medio de técnicas lineales y no lineales. El análisis en componentes principales (PCA, por sus siglas en inglés), análisis discriminante de Fisher y proyecciones aleatorias son ejemplos de técnicas lineales de reducción de dimensionalidad lineal que se han generalizado a técnicas no lineales por medio de métodos basados en kernels<sup>17</sup>. El análisis no lineal de variedades incluye metodologías tales como el encaje local lineal (LLE, por sus siglas en inglés), Isomap, y el encaje no lineal estocástico con distribución T. Estos métodos no lineales se desarrollaron para visualizar datos de dimensión muy alta por medio de un mapeo de encaje que los proyec-

ta a un subespacio de baja dimensionalidad<sup>21</sup>. También se han desarrollado métodos basados en ciertas suposiciones de independencia condicional y variables latentes que permiten calcular representaciones compactas de los datos. Estas técnicas incluyen el mapeo topológico, análisis de componentes independientes, análisis semántico probabilístico de variables latentes y la asignación latente de Dirichlet (LDA, por sus siglas en inglés)<sup>22</sup>. La metodología de reducción de dimensionalidad que se utiliza en este trabajo está basada en el algoritmo LDA, cuyo desempeño en la tarea de clasificación de cráneos sinostóticos es superior a la de otros algoritmos estándar que incluyen PCA, proyecciones aleatorias y métodos del estado del arte tales como el LLE e Isomap<sup>10</sup>.

La asignación latente de Dirichlet<sup>22</sup> es un modelo generativo Bayesiano para el análisis de datos de concurrencia que ha sido utilizado de manera muy exitosa en aplicaciones variadas que incluyen a los sistemas automáticos de recuperación de información (texto e imágenes). El modelo LDA trata a los datos como una colección calculada a partir de una mezcla, en la que los componentes de la mezcla se denominan *tópicos*<sup>23</sup>. Por ejemplo, si los datos son palabras contenidas en varios documentos, el modelo propone que cada documento es una mezcla de un pequeño grupo de *tópicos*, y que cada palabra ocurre con una frecuencia específica de acuerdo al *tópico* que la generó.

En este trabajo, el LDA se utiliza como un modelo generativo que muestrea *rasgos craneales de forma* a partir de una mezcla de *tópicos geométricos*. Un *tópico geométrico* se define como una distribución multinomial de las palabras en un *diccionario de rasgos craneales de forma*. La suposición es que las proporciones de la mezcla que están asociadas a cada una de las clases de cráneos (sagital, metópica, unicoronal) difieren significativamente, a pesar de que las instancias de las diferentes clases pueden tener rasgos craneales de forma en común, que son generados por el mismo *tópico geométrico*. El modelo LDA reduce la dimensionalidad utilizando una función que transforma las concurrencias de rasgos craneales de forma de un cráneo dado a un vector que almacena las proporciones correspondientes de la mezcla<sup>24</sup>. Los detalles del modelo LDA en el contexto de este trabajo se presentan en la sección método de reducción de dimensionalidad.

## PROPUESTA METODOLÓGICA

En la sección introducción se mencionó que una clase de cráneos puede caracterizarse en térmi-

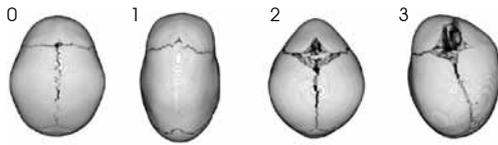
nos de patrones de concurrencias de *RCF*. Clases diferentes pueden tener rasgos craneales de forma en común. No obstante, es necesario que cada clase posea patrones de concurrencia distintivos para que sea posible clasificar con tasas de error razonables. La forma en que dichos patrones se organizan localmente origina algunas de las propiedades geométricas globales de una clase de cráneos.

La metodología que se propone en este trabajo consiste en una etapa de entrenamiento y una de prueba. La etapa de entrenamiento calcula una función de clasificación que se utiliza para predecir la clase de un cráneo que no ha sido visto con anterioridad. Para lograr este objetivo, el algoritmo de entrenamiento calcula patrones de concurrencia de *RCFs* a partir de un conjunto de entrenamiento. Este conjunto consiste en una muestra aleatoria de instancias de cráneos afectados por las varias clases de malformaciones causadas por craneosinostosis. Los patrones se calculan a partir de todas las instancias observadas en el conjunto de entrenamiento utilizando una *representación simbólica de contornos de forma*. Los patrones distintivos de cada clase de formas se calculan a partir de contornos simbólicos de forma por medio de un método de selección de variables basado en una prueba estadística de permutación<sup>24</sup> y un método estándar de reducción de dimensionalidad (LDA). La selección de variables y reducción de dimensionalidad *producen descriptores de forma* (DF) que son apropiados para entrenar a un clasificador.

Es importante mencionar que el conjunto de entrenamiento se divide en dos subconjuntos. Uno de ellos se utiliza para determinar qué patrones de concurrencia son distintivos (es decir, que separan las diversas clases de forma). El otro conjunto se utiliza para estimar los parámetros de una función de clasificación que minimiza el criterio de error *deja-un-dato-fuera* (LOO por sus siglas en inglés)<sup>17</sup>. La separación del conjunto de prueba en dos subconjuntos evita que el proceso de selección de rasgos introduzca un sesgo en el desempeño del clasificador. La etapa de prueba utiliza el clasificador y los patrones de concurrencia de *RCFs* de un cráneo nunca antes visto para predecir la clase a la que pertenece.

## Problemática

El problema específico a resolver es el siguiente: Se tiene una muestra aleatoria de  $N$  cráneos eti-



**Figura 2.** Reconstrucciones volumétricas del cráneo de un paciente sano (0) y pacientes afectados por craneosinostosis sagital (1), metópica (2) y unicoronal (3).

quetados como control (0), sagital (1), metópico (2), o unicoronal (3). Utilizando información de forma calculada a partir de volúmenes de TC y las etiquetas, se desea construir un conjunto de DF de baja dimensionalidad (DF de dimensión  $d < N$ ) y una función de clasificación que permita predecir la etiqueta de un cráneo nuevo. Ejemplos típicos de cráneos en la base de datos se muestran en la Figura 2.

### Algoritmo de entrenamiento

El objetivo del algoritmo de entrenamiento es calcular una función de clasificación capaz de predecir las etiquetas de formas craneales nunca antes vistas. Para simplificar la presentación del algoritmo de entrenamiento se consideran solamente dos clases de formas. La generalización del algoritmo para más de dos clases se presenta al final de esta sección. El algoritmo está dividido en ocho pasos que se resumen a continuación.

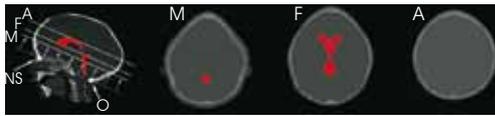
1. El primer paso extrae *contornos numéricos de forma* a partir de volúmenes de tomografía computarizada de un conjunto de entrenamiento.
2. El segundo paso calcula *contornos simbólicos de forma* etiquetando los vértices de los contornos numéricos de forma del primer paso.
3. El tercer paso calcula un diccionario de *rasgos craneales de forma (F)* aplicando un *algoritmo de bolsa de rasgos* a los contornos simbólicos del segundo paso.
4. El cuarto paso colecta concurrencias de rasgos craneales de forma en una matriz de concurrencia de tamaño  $N_c \times |F|$ .  $N_c$  es el número de clases de forma y  $|F|$  es el número de rasgos craneales de forma contenidos en el diccionario  $F$ . El  $i$ -ésimo renglón de la matriz de concurrencia es un vector de  $|F|$  dimensiones llamado *vector de patrón de concurrencias poblacionales*, cuyos componentes almacenan las frecuencias de concurrencia de los rasgos craneales de forma para la  $i$ -ésima clase de cráneos.

5. El quinto paso lleva a cabo un *proceso de selección de variables* respecto a los componentes de los vectores de concurrencia de patrones poblacionales. Esta tarea se ejecuta por medio de una prueba estadística de permutación que determina cuáles componentes muestran diferencias significativas entre las diferentes clases de forma. La salida de este paso es un vector binario de  $|F|$  dimensiones (llamado máscara de selección de rasgos) que indica cuáles componentes muestran diferencias significativas.
6. El sexto paso calcula *vectores de concurrencia de patrones individuales* que almacenan las frecuencias de concurrencia de rasgos craneales de forma para cada uno de los cráneos en el conjunto de entrenamiento. Estos vectores son utilizados para calcular vectores reducidos, que solamente incluyen aquellos componentes de los vectores de concurrencia de patrones individuales que son estadísticamente significativos, es decir, aquellos que fueron seleccionados en el quinto paso.
7. El séptimo paso disminuye la dimensionalidad de los vectores reducidos de concurrencias de patrones por medio de un método estándar de reducción de dimensionalidad. La salida de este paso son vectores de  $d$  dimensiones denominados DF.
8. Finalmente, el octavo paso utiliza los DF del séptimo paso (un vector para cada cráneo en el conjunto de entrenamiento), las etiquetas asociadas correspondientes, y una técnica de validación cruzada denominada *deja-uno-dato-fuera* con el objeto de encontrar los parámetros de un clasificador que se utiliza para predecir las etiquetas de cráneos nunca vistos.

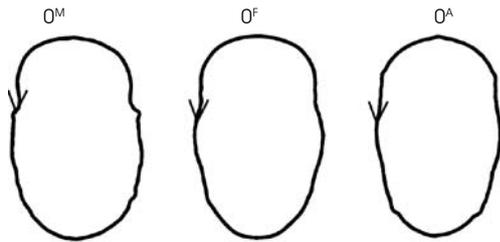
Los detalles del algoritmo de entrenamiento son como sigue: La entrada es un conjunto de volúmenes de TC,  $\{TC_1, \dots, TC_N\}$  y sus etiquetas correspondientes  $\{1, \dots, IN\}$ . Asumimos que existen dos clases de formas de cráneo (es decir,  $I_N \in \{0, 1\}$ ). La generalización del algoritmo a más de dos clases de formas es simple. La salida del algoritmo de entrenamiento es una función de clasificación  $\Psi$ .

### Contornos numéricos de forma

Los parámetros de entrada a este paso son  $\{TC_1, \dots, TC_N\}$  y los números  $N_v$  y  $N_l$  que definimos a continuación. Los contornos numéricos de forma se calculan a partir de imágenes de volúmenes de TC. La ubicación de los  $N_l$  planos de las imágenes en cada volumen están determinados por



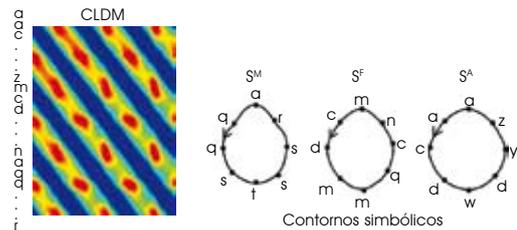
**Figura 3.** Los DF utilizados en este trabajo se calculan a partir de imágenes de TC calculadas en tres planos. Los planos son paralelos al plano que define la base del cráneo, que contiene a la recta definida por la sutura nasal (NS) y el opisthion (O). El plano M está localizado en el nivel del diámetro máximo del cuarto ventrículo, el plano F está localizado a la altura de la foramina de Munro, y el plano A sobre la superficie del ventrículo lateral. Los ventrículos se muestran en las imágenes como regiones sombreadas en rojo. Las imágenes fueron extraídas de un volumen de TC correspondiente a un paciente afectado por craneosinostosis metópica.



**Figura 4.** Contornos numéricos. Formas del cráneo de un sujeto afectado por craneosinostosis sagital. Los contornos orientados de forma en los planos M, F y A se calculan por medio de técnicas de segmentación e interpolación.

referencias cerebrales anatómicas bien definidas. Los planos son paralelos al plano especificado por la base del cráneo, que contiene a la recta definida por la sutura nasal y el opisthion. El plano M está localizado al nivel del diámetro máximo del cuarto ventrículo, el plano F está localizado a la altura de la foramina de Munro, y el plano A sobre la superficie del ventrículo lateral (Figura 3). En este caso tenemos  $N_L = 3$ . Los estudios de TC de la población bajo consideración consisten en volúmenes de 150 imágenes adquiridas con una resolución axial de 5 mm.

Los contornos orientados de forma asociados a los planos M, F y A, se calculan utilizando métodos estándar de segmentación e interpolación como se muestra en la Figura 4. Cada contorno se representa como un conjunto que posee  $N_V$  vértices de coordenadas  $\{x[n], n = 0, 1, \dots, N_V - 1\}$ . La colección de todos los contornos numéricos asociados con los volúmenes en el conjunto de entrenamiento se denota como  $O = \{O_1, \dots, O_N\}$ , en donde  $O_i = \{O_i^M, O_i^F, O_i^A\}$  y donde  $O_i^M, O_i^F, O_i^A$  son los contornos numéricos calculados en los planos M, F y A, respectivamente.



**Figura 5.** Los contornos simbólicos de forma se calculan etiquetando los vértices de los contornos numéricos. Las etiquetas se calculan por medio de un algoritmo de aglomeración ( $k$ -medias) aplicado a los CLDs asociados con los vértices de los contornos numéricos en el conjunto de entrenamiento  $O$ . Las columnas de la CLDM de la figura están etiquetadas con las letras 'aqqstsr' 'mcdmmqcn' 'aacdwdyz'. Estas etiquetas se asignan a los vértices correspondientes de los contornos numéricos. Todos los vértices que pertenecen al mismo grupo comparten la misma etiqueta.

### Contornos simbólicos de forma

La entrada de este paso es el conjunto de contornos numéricos  $O$  y el número  $k$  que se define a continuación. Los contornos simbólicos de forma se calculan etiquetando los vértices de los contornos numéricos de forma en  $O$  (Figura 5). Las etiquetas se calculan aplicando un algoritmo de aglomeración ( $k$ -medias) a un conjunto de atributos asociados a los vértices de los contornos numéricos de forma. El número  $k$  de grupos calculados por el algoritmo de aglomeración es definido por el usuario. En este trabajo utilizamos distribuciones de longitudes de cuerda como atributos numéricos de forma<sup>18</sup>.

Una medida de longitud de cuerda  $l_{ij}$  perteneciente a un contorno se define como la longitud del segmento de línea que une el par de vértices  $i$  y  $j$ , y está normalizada con respecto a una longitud característica (como la longitud de la cabeza). El conjunto completo de todas las cuerdas, está formado por las cuerdas dibujadas de cada vértice del contorno a los vértices restantes. El conjunto de longitudes de cuerda se puede representar por medio de una matriz de distancias que se denomina *matriz de distribución de longitud de cuerdas* (CLDM por sus siglas en inglés)<sup>18</sup>.

Debido a que los vértices de un contorno orientado pueden enumerarse en forma consecutiva, el  $i$ -ésimo renglón de la CLDM almacena las longitudes de cuerda del  $i$ -ésimo vértice a los vértices restantes. Al vector formado por el  $i$ -ésimo renglón de la CLDM se le denomina *distribución de longitud de cuerdas* (CDL por sus siglas en inglés). Con el objeto de evitar ambigüedades en la forma de numerar los vértices del contorno, el

primer vértice se define en términos de la ubicación de la sutura metópica del contorno. Debido a que la CDLM puede incorporar información de  $N_L$  contornos por medio del cálculo de distancias inter contorno e intra contorno, la matriz es en general de tamaño  $N_{LV} \times N_{LV}$  en donde  $N_{LV} = N_L N_V$ . Nótese que la CDLM contiene  $N_{LV}(N_{LV}-1)/2$  parámetros libres que son independientes del sistema coordinado en el que se encuentran representados los contornos.

La Figura 5 muestra los contornos simbólicos  $S^M$ ,  $S^F$  y  $S^A$  de los contornos de un cráneo y su correspondiente CDLM representada como una imagen en falso color. Cada vértice del contorno es etiquetado de acuerdo al grupo al que pertenece. Los parámetros de salida de este paso son un conjunto de contornos simbólicos de forma y una matriz  $C$  de tamaño  $k \times N_{LV}$  que almacena los centros asociados a los grupos calculados por el algoritmo de aglomeración. El conjunto de contornos simbólicos se denota como  $S = \{S_1, \dots, S_N\}$ , en donde  $S_i = \{S_i^M, S_i^F, S_i^A\}$ . Nótese que cada uno de los elementos de  $S_i$  consiste en un arreglo  $(x, y)$  de coordenadas y un arreglo que almacena las etiquetas correspondientes.

El conjunto  $S$  se divide en dos subconjuntos  $S_F$  y  $S_C$  ( $S_F \cap S_C = \emptyset$ ). Estos subconjuntos se obtienen por medio de muestreo sin reemplazo del conjunto  $S$ . Es conveniente mencionar que  $S_F$  se utiliza en los pasos 3-5 del algoritmo de entrenamiento y  $S_C$  en los pasos 6-8, para entrenar a la función de clasificación. La subdivisión de  $S$  evita que el proceso de selección de variables sesgue el desempeño de la función de clasificación.

### Diccionario de rasgos craneales de forma

El parámetro de entrada a este paso es el conjunto de contornos simbólicos de forma  $S_F$ . Las etiquetas asociadas a los contornos simbólicos de forma pueden utilizarse para construir  $RCF$ . El tamaño de un  $RCF$  es un número entero fijo en el intervalo  $1 \leq N_s \leq N_V$ . Por ejemplo, si  $N_s = 3$ , cada  $RCF$  está formado por tres etiquetas. Un ejemplo sencillo con  $N_L = 3$  ilustra el proceso de construcción de  $RCFs$  utilizando el algoritmo de bolsa de rasgos. Sean 'aacdwdyz', 'mcdmmqcn', 'aaqstsr' las cadenas de caracteres asociadas con los contornos simbólicos de forma  $S^A$ ,  $S^F$  y  $S^M$  mostrados en la Figura 5, en donde  $N_V = 8$ . Con  $N_s = 3$  los  $RCFs$  son: {'aac', 'acd', 'cdw', 'dwd', 'wdy', 'dyz', 'yza', 'zad', 'mcd', 'cdm', 'dmm', 'mmq', 'mqc', 'qcn', 'cnm', 'nmc', 'aaq', 'qqs', 'qst', 'sts', 'tss', 'ssr', 'sra', 'raq'}.

Los parámetros de salida de este paso son:  $A = \{a_1, \dots, a_{|A|}\}$ , los  $RCFs$  asociados con la clase 0 de formas craneales,  $B = \{b_1, \dots, b_{|B|}\}$ , el conjunto de todas las  $RCFs$  asociadas con la clase 1, y la colección de todos los *distintos*  $RCFs$  contenidos en  $A \cup B$ . Esta colección se conoce con el nombre del diccionario de rasgos craneales y se denota como  $F = \{f_1, \dots, f_{|F|}\}$ .

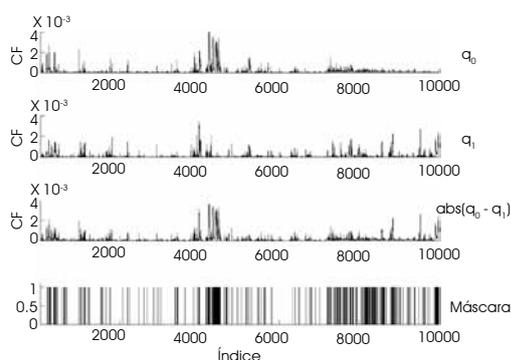
### Matriz de concurrencias de rasgos craneales

En este paso los rasgos de formas craneales contenidos en los conjuntos  $A$  y  $B$  son procesados con el objeto de calcular *concurrencias* de rasgos craneales de forma asociados con las clases 0 y 1. Las frecuencias de concurrencia se almacenan en una matriz  $Q$  de tamaño  $N_C \times |F|$ , en donde  $N_C$  es el número de clases de forma. El  $i$ -ésimo renglón de la matriz es un *vector de población de patrones de concurrencia* de  $|F|$  dimensiones denominado  $q_i$ . La  $(i, j)$ -ésima componente  $q_{ij}$  de  $q_i$  es un número que representa la frecuencia de concurrencia de la  $j$ -ésima forma craneal ( $f_j \in F$ ) de la  $i$ -ésima clase de formas.

### Selección de rasgos craneales de forma

Los parámetros de entrada a este paso son: la matriz de concurrencia  $Q$ , los rasgos craneales de forma de los conjuntos  $A$ ,  $B$ , y un nivel de significancia  $\alpha$ . El proceso de selección de rasgos consiste en dos etapas. En la etapa uno se calcula el *vector de diferencia de concurrencias*  $\mathbf{d} = \text{abs}(\mathbf{q}_0 - \mathbf{q}_1)$ , cuyo  $k$ -ésima componente está definido como  $d_k = \text{abs}(q_{0,k} - q_{1,k})$ , con  $k = 1, \dots, |F|$ . Este componente almacena el valor absoluto de la diferencia entre las frecuencias de concurrencia del  $k$ -ésimo rasgo de forma craneal entre las clases 0 y 1. En la segunda etapa, se utiliza una prueba estadística basada en permutaciones que es aplicada individualmente a cada componente del vector de diferencia de concurrencias con el objeto de determinar si las diferencias son significativas<sup>24</sup>. El objetivo de la prueba de permutación es determinar si es la ocurrencia de la diferencia medida (digamos  $d_k$ ) es un evento raro bajo la hipótesis nula, esto es, la hipótesis de que no hay diferencias significativas entre clases respecto a la frecuencia de concurrencia del  $k$ -ésimo  $RCF$ . Con el objeto de lograr este objetivo, el algoritmo calcula la distribución acumulativa empírica asociada con la hipótesis nula para cada uno de los componentes del vector  $\mathbf{d}$ .

Los parámetros de salida de este paso son: la máscara de selección de variables o vector de



**Figura 6.** El proceso de selección de rasgos consiste en dos etapas. a) el cálculo del vector de diferencia de concurrencias  $\mathbf{d}$  entre los vectores de patrones de concurrencia  $\mathbf{q}_0$  y  $\mathbf{q}_1$  asociados con las clases 0 y 1, respectivamente; y b) la determinación por medio de una prueba estadística de permutación de aquellos componentes del vector de diferencias que son significativas. El vector de diferencias está definido como  $\mathbf{d} = \text{abs}(\mathbf{q}_0 - \mathbf{q}_1)$ . La salida del proceso de selección de variables es una máscara binaria cuyos componentes están definidos como  $m_k = 1$  si la  $k$ -ésima componente de  $\mathbf{d}$ ,  $d_k$ , es estadísticamente significativa, y  $m_k = 0$  de otro modo. La figura muestra los vectores de patrones de concurrencia para las clases de cráneos metópicos ( $\mathbf{q}_0$ ) y normales ( $\mathbf{q}_1$ ). Los vectores de patrones de concurrencia poblacionales contienen  $|F| = 10,000$  componentes. En este ejemplo,  $|F^*| = 704$ . El gráfico en la parte de abajo de la figura muestra la máscara de selección de variables representada en la forma de un código de barras. Los gráficos se calcularon con  $N_V = 200$ ,  $N_L = 3$ ,  $k = 175$  y  $N_S = 9$ . El eje vertical corresponde a la frecuencia de concurrencias (CF, por sus siglas en inglés).

$|F|$  dimensiones  $\mathbf{m}$ , cuyos componentes están definidos como  $m_k = 1$  si la  $k$ -ésima componente de  $\mathbf{d}$ ,  $d_k$ , es estadísticamente significativa, y  $m_k = 0$  de otro modo. Se denota como  $|F^*|$  el número de elementos de  $\mathbf{m}$  diferentes de cero, es decir, el número de rasgos de forma craneal del denominado *diccionario reducido de rasgos de forma craneales*  $F^*$ . La Figura 6 muestra ejemplos de vectores de patrones de concurrencia, su vector de diferencias correspondiente y la máscara de selección de variables. Los detalles del algoritmo de permutación se detallan a continuación.

Los parámetros de entrada al algoritmo de permutación son: el nivel de significancia  $\alpha$ , el diccionario de rasgos craneales de forma  $F$ , el vector de diferencia de concurrencias  $\mathbf{d}$ , y los conjuntos de rasgos craneales de forma  $A = \{a_1, \dots, a_{|A|}\}$  y  $B = \{b_1, \dots, b_{|B|}\}$  asociados a las clases 0 y 1, respectivamente. La salida es un vector binario  $\mathbf{m} = [m_1, \dots, m_{|NF|}]$  que señala cuáles de los componentes del vector de diferencias son estadísticamente significativos al nivel de significancia  $\alpha$ .

1. Inicializar los componentes de  $\mathbf{m}$  con cero.
2. Calcular el conjunto  $C$  concatenando a los conjuntos  $A$  y  $B$ . Esto es,  $C = [A, B] = \{a_1, \dots, a_{|A|}, b_1, \dots, b_{|B|}\}$ .
3. Reordenar los elementos de  $C$  por medio de una permutación aleatoria. Sean  $C^* = [A^*, B^*] = \{a_1^*, \dots, a_{|A|}^*, b_1^*, \dots, b_{|B|}^*\}$  las muestras del arreglo permutado.
4. Calcular la matriz de concurrencia  $\mathbf{Q}^*$  a partir de  $A^*$  y  $B^*$ . Calcular el vector de diferencias  $\mathbf{d}^*$ .
5. Repetir los pasos 3 y 4,  $T$  veces para obtener la matriz  $\mathbf{D}^* = [\mathbf{d}_1^*, \dots, \mathbf{d}_T^*]$  de dimensiones  $N_{|F|} \times T$ .
6. Para cada  $j \in \{1, \dots, N_{|F|}\}$ , reordenar en forma creciente los elementos del  $j$ -ésimo renglón  $[d_{j1}^*, \dots, d_{jT}^*]$  de la matriz  $\mathbf{D}^*$  y construir la distribución acumulativa empírica. Rechazar la hipótesis nula respecto a la igualdad de frecuencia de concurrencias para el  $j$ -ésimo rasgo craneal de forma entre las clases 0 y 1 si  $d_{jj} \in \mathbf{d}$  está localizada por arriba del percentil  $(1 - \alpha)$ . Si la hipótesis nula se rechaza, entonces  $m_j = 1$ .

#### Cálculo de los vectores de patrones de concurrencia y los descriptores de forma\*

Los parámetros de entrada a este paso son: la máscara de selección de variables  $\mathbf{m}$ , el diccionario de rasgos craneales de forma  $F$ , el parámetro de dimensión  $d$  y los contornos simbólicos de forma contenidos en el conjunto  $S_C$ . Se aplica el algoritmo de bolsas de rasgos de la sección diccionario de rasgos craneales de forma a los elementos de  $S_C$  con el objeto de calcular el conjunto de rasgos craneales de forma para cada uno de los pacientes representados en este conjunto. El conjunto de rasgos craneales de forma asociados con el  $i$ -ésimo paciente en  $S_C$  se denota como  $RCF_i$ . Utilizando el diccionario  $F$  como referencia, se construye el vector de patrones de concurrencia  $\mathbf{u} = [u_1, \dots, u_{|F|}]$ , para cada uno de los elementos en el conjunto  $\{RCF_1, \dots, RCF_{|S_C|}\}$ . El número de componentes del vector  $\mathbf{u}$  se reduce descartando aquellos componentes para los que el procedimiento de selección de variables no detectó diferencias significativas, es decir aquellos componentes para los que  $m_k = 0$ . El resultado es un *vector reducido*  $\mathbf{r}$  cuya dimensionalidad es disminuida a un valor  $d$  por medio de un método de reducción de dimensionalidad. El parámetro de salida de este paso es un conjunto de vectores de  $d$  dimensiones denominados  $\mathbf{DF} = \{\mathbf{v}_1, \dots, \mathbf{v}_{|V|}\}$ , en donde  $|V| = |S_C|$  y  $\mathbf{v}_j = [v_{j1}, \dots, v_{jd}]$ .

**Cuadro 1.** Funciones de Kernel.

Kernel	Forma funcional
1. Lineal	$k_l(x, x') = \sum_i x x'_i$
2. Gaussiano	$k_G(x, x') = \exp\{-\gamma \ x - x'\ ^2\}$
3. Hellinger	$k_H(x, x') = \sum_i  (x_i)^{1/2} - (x'_i)^{1/2} $
4. Hellinger modificado	$k_{H2}(x, x') = \sum_i  (x_i)^{1/2} - (x'_i)^{1/2} ^2$
5. Variación total	$k_{VT}(x, x') = \sum_i  x_i - x'_i $
6. Chi cuadrada ( $\chi^2$ )	$k_{\chi^2}(x, x') = \sum_i  x_i - x'_i   x_i + x'_i ^{-1}$
7. Gaussiano modificado	$k_{GJ}(x, x') = \exp\{-k_j(x, x')\}$ con $J \in \{H, H2, VT, \chi^2\}$
8. Polinomial	$k_p(x, x') = (1 + \sum_i x x'_i)^n$ con $n$ entero.

### Función de clasificación

Los parámetros de entrada a este paso son los DF  $V$ , sus correspondientes etiquetas de clase, y un parámetro  $C$  cuyo significado se define a continuación. La salida de este paso es una función de clasificación  $\Psi$ . La función de clasificación es una máquina de vectores de soporte tipo  $\mathcal{E}$  ( $\mathcal{E}$ -SVM por sus siglas en inglés)<sup>17</sup>. Las máquinas de soporte vectorial son clasificadores que mapean patrones de entrada a un espacio de dimensionalidad alta (posiblemente infinita) llamado *espacio de rasgos* o *características*. En este espacio construyen un hiperplano de margen máximo que separa las instancias de clases de objetos para predecir la etiqueta de una instancia nueva. El truco matemático que hace efectivas a las máquinas de soporte vectorial es que el cálculo del hiperplano en el espacio de características sólo depende de la llamada función de Kernel<sup>17,25</sup>. En este trabajo las funciones de Kernel están listadas en el Cuadro 1.

El entrenamiento de la función de clasificación se lleva a cabo minimizando un criterio de validación cruzada denominado *deja-un-dato-fuera* que minimiza el valor esperado del error de clasificación<sup>18</sup>.

La idea básica del criterio LOO es calcular un clasificador utilizando  $m-1$  patrones, dejando el  $i$ -ésimo patrón fuera y, subsecuentemente, calcular el error de predicción. El error total es el producto del promedio de todos los  $m$  posibles patrones.

### Algoritmo de prueba

Los parámetros de entrada al algoritmo de prueba son: el volumen de  $TC$  de un paciente nuevo  $TC_{nuevo}$ , la matriz de centros de grupos  $C$ , el diccionario de rasgos craneales de forma  $F$ , la máscara de selección de variables  $m$ , y la función de clasificación

$\Psi$ . La salida del algoritmo es la etiqueta correspondiente a la nueva forma craneal  $I_{nuevo}$ .

1. Calcular los contornos numéricos  $O_{nuevo}$  a partir del volumen  $TC_{nuevo}$ .
2. Calcular los contornos simbólicos de forma  $S_{nuevo}$  etiquetando los vértices de  $O_{nuevo}$ . Las etiquetas se calculan utilizando una regla de *vecino más cercano* que utiliza a la matriz de centros  $C$  como referencia.
3. Calcular el conjunto de rasgos craneales de forma a partir de los contornos simbólicos  $S_{nuevo}$  utilizando el algoritmo de bolsa de rasgos descrito en el tercer paso del algoritmo de entrenamiento. Descartar todos aquellos rasgos craneales de forma que no aparecen en el diccionario  $F$ . Calcular el conjunto de rasgos craneales de forma  $RCF_{nuevo}$  y calcular el descriptor de forma  $v_{nuevo}$  a partir de la aplicación de la máscara de selección de rasgos al vector  $u_{nuevo}$ , y la aplicación del método de reducción de dimensionalidad al vector  $r_{nuevo}$  (ver sección cálculo de los vectores de patrones de concurrencia y los descriptores de forma)
4. Calcular  $I_{nuevo} = \Psi(v_{nuevo})$ .

### Algoritmos de entrenamiento y prueba en el caso de clases craneales múltiples

El algoritmo de entrenamiento descrito en la sección algoritmo de entrenamiento asume la existencia de dos clases de formas craneales. Cuando hay más de dos, el algoritmo de selección de variables (sección Selección de rasgos craneales de forma) y la función de clasificación necesitan modificarse como sigue:

5. En el caso binario, el algoritmo de selección de rasgos se encarga de determinar cuáles componentes del vector de diferencias  $d$ , revelan diferencias significativas entre las clases 0 y 1. La salida del algoritmo es una máscara binaria que indica cuáles componentes son significativos. En el caso de que haya  $N_C$  clases, se deben de calcular un total de  $R = N_C(N_C - 1)/2$  vectores correspondientes a los  $R$  pares diferentes. En este caso, la salida del algoritmo de selección de rasgos es una máscara binaria calculada como el producto componente a componente de las máscaras calculadas para cada par de clases. El producto se define como la operación booleana «O». Nótese que esta modificación considera significativos a aquellos rasgos craneales de

forma que fueron señalados como significativos en al menos una las  $R$  pruebas llevadas a cabo con pares de clases.

- En el caso de clases múltiples, la función de clasificación utiliza un *esquema de conteo de uno contra todos*<sup>18</sup> con el objeto de determinar la clase de una nueva forma craneal. En este método se calcula un clasificador binario para cada par de clases. La clase que resulta con más votos determina la etiqueta asignada a un nuevo patrón.

### Método de reducción de dimensionalidad

En esta sección se describe brevemente el modelo probabilístico utilizado para disminuir la dimensionalidad de los vectores de patrones de concurrencia reducidos de la sección cálculo de los vectores de patrones de concurrencia y los descriptores de forma. Los detalles técnicos de esta metodología pueden encontrarse en las referencias<sup>22,23</sup>.

El modelo LDA es más fácil de comprender si se analiza el proceso de generación de los vectores de patrones de concurrencia reducidos que se asume en este modelo. Una primera suposición del modelo es que existen  $d$  tópicos geométricos (ver sección métodos de reducción de dimensionalidad). Cada tópico se define como una distribución multinomial sobre el vocabulario de tamaño  $|F|$ . Para cada contorno simbólico  $S_j$ , primero se muestrean las proporciones de la mezcla  $\theta_j = \{\theta_{jk}\}$ , con respecto a los  $d$  tópicos geométricos utilizando una distribución simétrica de Dirichlet caracterizada por un parámetro de concentración  $\gamma$ . Para el  $i$ -ésimo rasgo craneal de forma  $RCF_i$ , de  $S_j$  (que se denota como  $x_{ij}$ ) un tópico  $z_{ij}$  se muestrea con el tópico  $k$  escogido con probabilidad  $\theta_{jk}$ . Entonces, el rasgo craneal de forma  $x_{ij}$  se muestrea del tópico geométrico  $z_{ij}$ , con  $x_{ij}$  tomando el valor  $w$  con probabilidad  $\phi_{kw}$ . Finalmente, se utiliza una distribución a priori simétrica de Dirichlet con parámetro de concentración  $\beta$  para los parámetros de los tópicos  $\theta_k = \{\phi_{kw}\}$ .

Asumiendo que  $\gamma$  y  $\beta$  se conocen, la distribución de todos los parámetros y variables del modelo se escribe como:

$$p(x, z, \theta, \phi | \gamma, \beta; d) = \left[ \prod_{j=1}^D \frac{\Gamma(d\gamma)}{\Gamma(\gamma)^d} \prod_{k=1}^d \theta_{jk}^{n_{jk} + \gamma - 1} \right] \left[ \prod_{k=1}^d \frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \prod_{w=1}^W \phi_{kw}^{n_{kw} + \beta - 1} \right]$$

en donde  $\mathbf{z} = \{z_{ij}\}$ ,  $\mathbf{x} = \{x_{ij}\}$ ,  $\theta = \{\theta_j\}$ ,  $\phi = \{\phi_j\}$ ,  $n_{jkw} = \#\{i: x_{ij} = w, z_{ij} = k\}$ ,  $\Gamma$  es la función gamma, y el punto significa sumar sobre el índice correspondiente:  $n_{\cdot kw} = \sum_i n_{ijkw}$  y  $n_{jk\cdot} = \sum_w n_{jkw}$ . Dados los rasgos craneales de forma  $\mathbf{x}$ , la tarea de la inferencia

Bayesiana es calcular la distribución posterior con respecto a los tópicos latentes geométricos  $\mathbf{z}$ , las proporciones de la mezcla  $\theta$ , y los parámetros de los tópicos geométricos  $\phi$ . La inferencia se puede llevar a cabo por medio del muestreo de Gibbs colapsado<sup>28</sup>. El resultado deseado es

$$p(z_{ij} = k | z^{-ij}, x, \gamma, \beta; d) = \frac{n_{\cdot kw}^{-ij} + \beta}{n_{\cdot k}^{-ij} + W\beta} \cdot \frac{n_{jk}^{-ij} + \gamma}{n_{j\cdot}^{-ij} + d\gamma}$$

en donde  $-ij$  significa que las variables  $z_{ij}$  y  $x_{ij}$  se excluyen de la expresión. Habiendo obtenido la distribución condicional completa, el algoritmo de Gibbs es como sigue: Las variables  $z_{ij}$  se inicializan con valores  $\{1, \dots, d\}$ , determinando de esta manera el estado inicial de la cadena de Markov. Entonces, la cadena se corre por un cierto número de iteraciones, calculando un nuevo estado  $z_{ij}$  muestreando a partir de la distribución condicional del párrafo anterior. Después de un cierto número de iteraciones, la cadena aproximará la distribución objetivo y se almacenan los valores de los estados  $z_{ij}$ . Los valores de  $\theta$  y  $\phi$  dados  $\mathbf{z}$  y  $\mathbf{x}$  pueden estimarse a partir de las muestras de la cadena como

$$\hat{\theta}_{jk} = \frac{n_{jk} + \gamma}{n_{j\cdot} + d\gamma}$$

$$\hat{\phi}_{kw} = \frac{n_{\cdot kw} + \beta}{n_{\cdot k} + W\beta}$$

Nótese que

$$p(z_{ij} = k | z^{-ij}, x) \propto (n_{\cdot kw}^{-ij} + \beta) (n_{j\cdot}^{-ij} + W\beta)^{-1} (n_{jk}^{-ij} + \gamma);$$

por esta razón,  $z_{ij}$  depende de  $z^{ij}$ , a través de las cuentas  $n_{\cdot kw}$ ,  $n_{\cdot k}$ ,  $n_{jk}$ . Específicamente, se observa que la convergencia del muestreador de Gibbs es rápida porque la dependencia de una variable  $z_{ij}$  con respecto a cualquier otra variable  $z_{lm}$  es débil para conjuntos de datos grandes<sup>24</sup>. Reducción de dimensionalidad con el LDA se alcanza transfor-

**Cuadro 2.** Valores de los parámetros utilizados en los experimentos.

Parámetro	Rango	Parámetro	Rango
$N_V$	200	$\alpha$	0.05
$N_L$	3	$ Sc $	0.75 $ S $
$N_C$	3	$ Sf $	0.25 $ S $
$N_s$	3,5,7,9,11	$\beta$	1,2,...,5
$k$	50,75,100, 125,...,250	$\gamma$	1,5,10,...,60
$d$	1,2,...,10	$n$	1,2,...,6

mando los vectores  $u_j$  a un vector de las proporciones de la mezcla:

$$u_j \mapsto (v_{j1}, \dots, v_{jd}) = (\hat{\theta}_{j1}, \dots, \hat{\theta}_{jd}).$$

### RESULTADOS

En esta sección presentamos los resultados de un número extenso de experimentos de clasificación. Nuestra base de datos contiene 102 cráneos con sinostosis sagital, 42 cráneos con sinostosis metópica, 12 cráneos con sinostosis sagital y 65 cráneos de individuos sanos (población de control). Los parámetros de entrada utilizados en los experimentos de clasificación se muestran en el Cuadro 2.

Los resultados de clasificación se muestran en los Cuadros 3 y 4. Se reportan resultados para los llamados DF modificados (DFM). Los DFM se calcularon sin considerar el algoritmo de selección de variables. Los experimentos con DFM permiten evaluar el efecto del algoritmo de selección de variables en el contexto del estudio comparativo con los kernels listados en el Cuadro 1.

La tasa de error de clasificación más baja se obtuvo con el Kernel  $k_{GH2}$  (3.3%) calculado con DF

**Cuadro 3.** Tasas de error de clasificación y de reducción de dimensionalidad (TRD) calculadas para los DFM.

	Error de clasificación		Error de clasificación	
	(%)	TRD	(%)	TRD
$k_L$	4.95	67:1	$k_{\chi^2}$	10.99 75:1
$k_G$	4.95	67:1	$k_{GH1}$	5.49 87.5:1
$k_{H1}$	8.24	120:1	$k_{GH2}$	4.95 67:1
$k_{H2}$	8.24	87.5:1	$k_{GTV}$	5.49 67:1
$k_{TV}$	7.69	100:1	$k_{GX^2}$	4.95 100:1
$k_P$	4.40	67:1	—	—

**Cuadro 4.** Tasas de error de clasificación y de reducción de dimensionalidad (TRD) calculadas para los DF.

	Error de clasificación		Error de clasificación	
	(%)	TRD	(%)	TRD
$k_L$	4.95	100:1	$k_{\chi^2}$	14.29 75:1
$k_G$	4.4	60:1	$k_{GH1}$	4.95 87.5:1
$k_{H1}$	8.79	87.5:1	<b><math>k_{GH2}</math></b>	<b>3.30 67:1</b>
$k_{H2}$	8.24	100:1	$k_{GTV}$	4.95 67:1
$k_{TV}$	8.24	100:1	$k_{GX^2}$	4.95 100:1
$k_P$	5.35	120:1	—	—

(Cuadro 4). La tasa de error más alta se obtuvo con el Kernel  $k^2$  (14.29%) también calculado con DF. El Kernel lineal alcanza una tasa de error de clasificación del 4.95% para los DF y DFM. El Kernel polinomial aplicado a DFM alcanza una tasa de 4.4%. El Kernel Gaussiano estándar alcanza la misma tasa de clasificación con DFs.

Los Cuadros 5 y 6 muestran las matrices de confusión para los experimentos de clasificación. Las tasas de clasificación condicional más altas se obtienen con el Kernel  $k_{GH2}$  aplicado a los descriptores DF. Las tasas de clasificación condicional representadas como  $(P_{11}, P_{2|2}, P_{0|0})$  son (98.9, 94.74, 94.34), respectivamente (Cuadro 6).

### DISCUSIÓN Y CONCLUSIONES

La CP es un área de investigación importante que requiere el desarrollo de nuevas metodologías para caracterizar el complejo craneofacial. Métodos cuantitativos para describir la forma del cráneo son fundamentales para el estudio de anomalías craneales y su relación con el estado fisiológico de los pacientes afectados. Un paso importante en esta dirección es el desarrollo de métodos efecti-

**Cuadro 5.** Matrices de confusión (%) para los experimentos con DFM. S, M y C denotan las clases sagital, metópica y control, respectivamente.

	$k_L$			$k_{\chi^2}$		
	S	M	C	S	M	C
S	96.7	2.22	1.1	95.6	4.4	0
M	0	94.74	5.26	2.63	84.21	13.16
C	5.66	1.89	92.45	0	9.43	90.57
S	96.7	$k_G$ 2.2	1.1	95.6	$k_{GH1}$ 1.1	3.3
M	0	94.74	5.26	0	94.74	5.26
C	3.77	3.77	92.45	3.77	3.77	92.45
S	92.31	$k_{H1}$ 2.2	5.49	96.7	$k_{GH2}$ 2.2	1.1
M	0	89.47	10.53	0	94.74	5.26
C	3.77	3.77	92.45	3.77	3.77	92.45
S	91.2	$k_{H2}$ 8.79	0	95.6	$k_{GTV}$ 2.2	2.2
M	0	92.11	7.89	0	94.74	5.26
C	1.89	5.66	92.45	5.66	1.89	92.45
S	93.41	$k_{TV}$ 2.2	4.4	95.6	$k_{GX^2}$ 2.2	2.2
M	0	89.47	10.53	0	94.74	5.26
C	1.89	5.66	92.45	3.77	3.77	92.45
S	96.7	$k_P$ 2.2	1.1	—	—	—
M	0	93.37	2.63	—	—	—
C	3.73	3.73	92.45	—	—	—

**Cuadro 6.** Matrices de confusión (%) para los experimentos con DF. S, M y C denotan las clases sagital, metópica y control, respectivamente.

	S	$k_L$ M	C	S	$k_{X^2}$ M	C
S	97.8	1.1	1.1	91.21	2.2	5.49
M	0	94.74	5.26	0	89.47	1.53
C	7.55	1.89	90.57	9.43	16.98	73.58
	S	$k_G$ M	C	S	$k_{GH1}$ M	C
S	98.9	1.1	0	97.8	2.2	0
M	0	86.84	13.16	0	92.11	7.89
C	1.89	1.89	96.23	5.66	3.77	90.50
	S	$k_{H1}$ M	C	S	$k_{GH2}$ M	C
S	93.41	2.2	4.4	98.9	1.1	0
M	0	86.84	13.16	0	94.74	5.26
C	3.77	5.66	90.57	3.77	1.89	94.34
	S	$k_{H2}$ M	C	S	$k_{GTV}$ M	C
S	95.6	2.2	2.2	96.7	1.1	2.2
M	0	86.84	13.16	2.63	92.11	5.26
C	3.7	7.55	88.68	3.77	0	96.23
	S	$k_{TV}$ M	C	S	$k_{GX^2}$ M	C
S	93.41	3.3	3.3	96.7	2.2	1.1
M	0	92.11	7.89	0	94.74	5.26
C	3.77	7.55	86.68	5.66	1.89	92.45
	S	$k_p$ M	C	S		C
S	98.9	1.1	0	—	—	—
M	0	97.37	26.3	—	—	—
C	5.66	9.43	84.91	—	—	—

vos para clasificar malformaciones causadas por craneosinostosis.

En este trabajo evaluamos una propuesta metodológica a través de un estudio comparativo en el que se midió el desempeño de clasificación de un clasificador lineal y varios clasificadores que utilizan Kernel no lineales listados en el Cuadro 1.

Dentro de los kernels no lineales se consideraron funciones estándar como  $v$ . el Kernel Gaussiano y el Kernel polinomial. Se utilizaron también kernels no lineales derivados de métricas Hilbertianas (kernels 3-6 en Cuadro 1). Estos últimos son útiles para cuantificar similitudes entre patrones que representan distribuciones discretas de probabilidad<sup>25</sup>. Tal es el caso de los DF generados por medio del modelo de reducción de dimensionalidad LDA. Se probaron además kernels no lineales generados a través de la composición de la función exponencial y los kernels Hilbertianos (Kernel 7 en Cuadro 1).

Los resultados obtenidos muestran que el desempeño del Kernel lineal (Cuadros 3 y 4) es del orden del 5%. Esto sugiere que los patrones sujetos a clasificación son aproximadamente separables en el espacio de características, que es esencial-

mente el mismo que el espacio original, dada la naturaleza del Kernel lineal. En general, los kernels lineales producen clasificadores que generalizan apropiadamente y tienen menor probabilidad de sobreajustar a los datos. Sin embargo, los kernels no lineales pueden refinar significativamente el proceso de clasificación, si se utilizan métodos de validación cruzada para minimizar la posibilidad de sobreajuste.

Los resultados de clasificación revelan que el Kernel no lineal que utiliza la distancia de Hellinger modificada alcanza una tasa de error de clasificación de 3.3%. Aproximadamente dos puntos porcentuales menor a la obtenida con el Kernel lineal. Este Kernel también supera las tasas de clasificación condicionales de todos los demás clasificadores que se probaron (Cuadros 5 y 6). Las tasas condicionales de error son una medida importante del desempeño de clasificación en experimentos que involucran varias clases de malformaciones<sup>21</sup>.

En un trabajo que se ha sometido a revisión<sup>10</sup> presentamos un estudio comparativo en el que se evalúa el desempeño de los kernels lineales y los DF con respecto al desempeño de descriptores numéricos tales como los descriptores de Fourier, el contexto de forma<sup>11</sup> y tecnologías basadas en morfometría y análisis euclidiano de matrices de distancia<sup>8,9</sup>. También se estudiaron métodos alternativos al LDA para reducir la dimensionalidad. Estos métodos incluyeron el PCA<sup>18</sup>, proyecciones aleatorias<sup>28</sup>, PLSA<sup>29</sup>, LLE<sup>30</sup>, e Isomap<sup>31</sup>, entre otros. Los resultados de aquel estudio sugirieron que la combinación de los DF y la técnica LDA resultan en las tasas más bajas de error de clasificación. Sin embargo, se consideró necesario probar kernels no lineales para determinar si era posible mejorar las tasas condicionales de clasificación. Los resultados obtenidos indican que el uso de clasificadores no lineales tiene el efecto deseado.

Los kernels no lineales pueden utilizarse para estudiar asociaciones entre los patrones observados (malformaciones craneales) y variables de interés (como el estado neurofisiológico de un paciente) a través de modelos de regresión. Por esta razón sugerimos que el Kernel de Hellinger modificado puede utilizarse para crear modelos predictivos, a partir de máquinas de vectores de soporte de regresión, en investigaciones en craneosinostosis. Los resultados de los Cuadros 2 y 3 también sugieren que el Kernel Gaussiano tiene un error de clasificación bajo, por lo que es factible utilizar métodos de regresión basados en procesos Gaussianos<sup>32</sup>.

Una dirección de trabajo futuro derivada de esta investigación consiste en probar el desempeño de clasificación de métodos que modelan en forma unificada la reducción de dimensionalidad y la clasificación en el contexto Bayesiano<sup>33-35</sup>. Finalmente, se hace notar que el algoritmo de selección de variables empleado se implementó por medio de una técnica de muestreo estocástico. También se hicieron pruebas con un método basado en bootstrap y se encontraron resultados comparables<sup>36</sup>.

En este trabajo hemos presentado una metodología para clasificar malformaciones craneales causadas por craneosinostosis simple. La metodología presentada está basada en el modelado estadístico de las concurrencias de rasgos craneales de forma. La metodología permite clasificación efectiva de formas craneales mediante la combinación de un proceso de selección de variables, una metodología de reducción de dimensionalidad y kernels no lineales. Para validar la metodología se realizaron una serie de experimentos de clasificación que mostraron que un Kernel no lineal basado en la distancia de Hellinger modificada alcanza una tasa de clasificación superior a las obtenidas en trabajos previamente publicados. La metodología y la evidencia experimental presentada en este estudio sugieren que la tecnología propuesta será útil para cuantificar malformaciones craneales en investigaciones relacionadas con la craneosinostosis primaria.

## REFERENCIAS

- Cohen MM, MacLean MC. *Craniosynostosis: Diagnosis, Evaluation, and Management*, 2ª Ed. Oxford University Press (Inglaterra), 2000.
- Shuper A, Merlob P, Grunembaum M, Reisner SH. The incidence of isolated craniosynostosis in the newborn infant. *Am J Dis Child* 1985; 139(1): 85-86.
- Speltz ML, Kapp-Simon KA, Cunningham ML, Marsh J, Dawson G. Single-suture craniosynostosis: a review of neurobehavioral research and theory. *Journal of Pediatric Psychology* 2004; 29(8): 651-668.
- Lajeunie E, Le Merrer M, Marchac C, Renier D. Genetic study of scaphocephaly". *Am. J Med Gene* 1996; 62: 282-285.
- Wikie AOM. *Craniosynostosis: genes and mechanisms*". *Human Molecular Genetics* 1997; 6(10): 1647-1656.
- Ruiz-Correa S, Sze RW, Starr JR, Lin HJ, Speltz ML, Cunningham ML et al. New scaphocephaly severity indices of sagittal craniosynostosis. A quantitative study with cranial index quantifications. *The American Cleft Palate-Craniofacial Association Journal*, 2006; 43(2): 211-221.
- Ruiz-Correa S, Starr JR, Lin HJ, Kapp-Simon K, Sze RW, Ellenbogen RG et al. New severity indices for quantifying single suture metopic craniosynostosis. *Neurosurgery* 2008; 63(2): 318-24; discussion 324-5.
- Bookstein FL. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991.
- Lele SR, Richtsmeier TJ. *An invariant approach to the statistical analysis of shapes*, New York: Chapman and Hall/CRC (EUA); 2001.
- Ruiz-Correa S, Marroquin JL, Gatica-Perez D, Lin HJ, Shapiro LG, Sze RW. Discriminating cranial malformations from CT imaging by co-occurrences of skull shape features. *Sometido a revisión*.
- Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts, *IEEE Transactions on PAMI*, 2002; (24)24: 509-522.
- Athitsos V, Wang J, Scarloff S, Betke M. Detecting instances of classes that exhibit variable structure'. *Proc. European Conference on Computer Vision*, 2006.
- Sebastian TB, Klein PN, Kimia BB. Recognition of shapes by editing shock graphs'. *Proc. International Conference of Computer Vision (ICCV)*, 2001: 755-762.
- Zhu SC, Yuille AL. FORMS: a flexible object recognition and modeling system'. *International Journal of Computer Vision* 1996; 20: 187-212.
- Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models - their training and application. *Computer Vision and Image Understanding* 1995; 61: 38-59.
- Captier G, Bigorre M, Rakotoarimanana JL, Leboucq N, Montoya P. Étude des variations morphologiques des scaphocéphalies. Implication pour leur systématisation. *Annales de Chirurgie Plastique et Esthétique* 2006; 50(6): 715-722.
- Scholkopf B, Somola A. *Learning wit kernels*. The MIT Press (EUA), 2002.
- Rangaraj RM. *Biomedical image analysis*. CRC Press (EUA), 2006.
- Ruiz-Correa S, Sze RW, Lin HJ, Shapiro LG, Speltz ML and Cunningham ML. Classifying craniosynostosis deformations from skull shape imaging, *Computer-Based Medical Systems (CBMS)*. *Proceedings of the 18th IEEE Symposium* 2005: 335-340.
- Ruiz-Correa S, Gatica-Perez D, Lin HJ, Shapiro LG, and Sze RW. A Bayesian Hierarchical Model for Classifying Craniofacial Malformations from CT Imaging. *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, Canada, 2008: 4063-4069.
- Lee JA, Verwey M. *Nonlinear dimensionality reduction*. Springer, New York, N.Y. (EUA), 2007.
- Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 2003: 993-1022.
- Griffiths TL, Steyvers M. Finding scientific topics. *Proceedings of the National Academy of Sciences* 2004; 101: 5228-5235.
- Golland P, Liang F, Mukherjee S, D. Panchenko D. Permutation test for classification. *COLT, LNAI*, 2005: 501-515.
- Hein M, Bousquet O. Hilbertian metrics and positive definite kernels on probability measures. *Proc. AISTATS* 2005.
- Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in practice*. London: Chapman and Hall, 2005.
- Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in practice*. London: Chapman and Hall (Inglaterra), 2005.
- Dasgupta S, Hsu DJ, Verma N. A concentration theorem for projections, *Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.
- Hofmann T. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning* 2001; 42: 177-196.

30. Roweis ST, Saul SK. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 2000; 290(5500): 2323-2326.
31. Tanenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000; 290(5500): 2319-2323.
32. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. MIT Press (EUA), 2006.
33. Fergus R, Perona P, Zisserman A. A visual category filter for Google images. In *Proc. ECCV*, Springer, 2004.
34. Cao L, Fei-Fei L. Spatially coherent latent topic model for concurrent object segmentation and classification. *IEEE Intern. Conf. in Computer Vision (ICCV)*. 2007.
35. Huang L, Yan D, Jordan MI. DiscLDA: Discriminative learning for dimensionally reduction and classification. *Advances in Neural Information Processing Systems (NIPS)*, 21, 2008.
36. Romano PJ. Bootstrap and randomization tests of some nonparametric hypothesis. *The Annals of Statistics* 1989; 17(1): 141-159.