

# Predicción del consumo de éxtasis a partir de redes neuronales artificiales

PALMER POL, A.\*; MONTAÑO MORENO, J.J.\*; CALAFAT FAR, A.\*\*

\* Facultad de Psicología. Universidad de las Islas Baleares.

\*\* IREFREA España

Enviar correspondencia a:

Alfonso Palmer Pol. Universidad de las Islas Baleares. Facultad de Psicología. Cra. de Valldemossa, km. 7.5. 07071 Palma de Mallorca (Baleares).

Teléfono 971173432; e-mail: alfonso.palmer@uib.es

## Resumen:

El propósito del presente estudio fue mostrar cómo una red neuronal artificial (RNA) puede ser útil para predecir el consumo de éxtasis (MDMA). Más específicamente, se trata de desarrollar una red neuronal del tipo *backpropagation* capaz de discriminar entre quién consume éxtasis y quién no, a partir de las respuestas dadas por los sujetos a un cuestionario. La muestra estaba compuesta por 148 consumidores y 148 no consumidores de éxtasis. Se explican las diferentes fases llevadas a cabo para desarrollar la RNA: selección de las variables relevantes y preprocesamiento de los datos, división de la muestra en grupo de entreno, validación y test, entreno y evaluación del modelo de red, y análisis de sensibilidad. La eficacia de la RNA entrenada fue del 96.66%. El área bajo la curva ROC (*Receiver operating characteristic*) fue de 0.99440.0055 SE. Por otra parte, se pretende mostrar que las RNA no representan una "caja negra", sino que pueden dar información acerca del grado de influencia que tiene cada variable predictora sobre el consumo de éxtasis.

**Palabras clave:** *redes neuronales artificiales, éxtasis, factores de riesgo, clasificación de patrones.*

## Abstract:

The purpose of this study was to show how an artificial neural network (ANN) can be useful to predict ecstasy (MDMA) consumption. More specifically, we tried to develop a backpropagation neural net capable to discriminate between who consumes ecstasy and who not, through the answers given by the subjects to a questionnaire. The sample was composed of 148 ecstasy consumers and 148 no consumers. We explain the different stages carried out to develop the ANN: selection of relevant variables and preprocessing of data, division of the sample into training, validation and test sets, training and evaluation of neural model, and sensitivity analysis. The accuracy of the ANN trained were 96.66%. The area under the ROC (Receiver operating characteristic) curve was 0.99440.0055 SE. On the other hand, we try to show that the ANN don't represent a "black box", but it can lead to useful insights into the roles played by different predictive variables in determining ecstasy consumption.

**Key words:** *artificial neural networks, ecstasy, risk factors, pattern classification.*

## INTRODUCCIÓN

Las Redes Neuronales Artificiales (RNA) son sistemas de procesamiento de la información cuya estructura y funcionamiento están inspirados en las redes neuronales biológicas (Hilera y Martínez, 1995). Consisten en un gran número de elementos simples de procesamiento llamados nodos o neuronas que están organizados en capas. Cada neurona está conectada con otras neuronas mediante enlaces de comunicación, cada uno de los cuales tiene asociado

un peso. En los pesos se encuentra el conocimiento que tiene la RNA acerca de un determinado problema.

La utilización de las RNA puede orientarse en dos direcciones, bien como modelos para el estudio del sistema nervioso y los fenómenos cognitivos, bien como herramientas para la resolución de problemas prácticos como la clasificación de patrones y la aproximación de funciones. Desde esta segunda perspectiva, las RNA han sido aplicadas de forma satisfactoria en la predicción de diversos problemas en diferentes áreas de conocimiento —biología, medicina, economía, ingenie-

ría, psicología, etc.— (Arbib, 1995; Simpson, 1995; Arbib, Erdi y Szentagothai, 1997); obteniendo excelentes resultados respecto a los modelos derivados de la estadística clásica (West, Brockett y Golden, 1997; De Lillo y Meraviglia, 1998; Jang, 1998; Waller, Kaiser, Illian et al., 1998). La virtud de las RNA reside en su capacidad para aprender funciones complejas o no lineales entre variables sin necesidad de imponer presupuestos o restricciones de partida en los datos.

El uso de esta tecnología computacional es relativamente reciente en el problema de las conductas adictivas (Palmer y Montaña, 1999). En este sentido, el Centro de Investigación Semeion de las Ciencias de la Comunicación (Roma, Italia), fundado y dirigido por Massimo Buscema, puede ser considerado como pionero en la aplicación de las RNA en este campo. Los investigadores de dicho centro han construido diferentes modelos de red con el fin de predecir el consumo de droga —sobre todo heroína— (Buscema, 1995; Buscema, Intraligi y Bricolo, 1998; Maurelli y Di Giulio, 1998; Speri, Schilirò, Bezzetto et al., 1998), extraer las características prototípicas del sujeto adicto (Buscema, Intraligi y Bricolo, 1998) y así, determinar el tratamiento más adecuado en función de esas características (Massini y Shabtay, 1998). Aunque los resultados son preliminares, estos trabajos demuestran que los buenos resultados obtenidos hasta el momento en las diferentes áreas de conocimiento se pueden extender al campo de las adicciones.

Siguiendo la línea de investigación iniciada por el equipo de Buscema, nos hemos propuesto llevar a cabo la aplicación práctica de una red neuronal para la predicción del consumo de éxtasis (MDMA) y determinar la influencia de cada variable predictora sobre este tipo de conducta. Más concretamente, se trata de construir un modelo de red neuronal que a partir de las respuestas de los sujetos a un cuestionario, sea capaz de discriminar entre quién consume éxtasis y quién no.

En este sentido, el consumo de éxtasis y otros derivados de las feniletilaminas ha experimentado un aumento significativo en los últimos años aunque más recientemente dicho uso ha experimentado una cierta estabilización o incluso descenso desde los niveles de consumo tan altos que había (Plan Nacional sobre Drogas, 2000). En la encuesta escolar, tras haber crecido espectacularmente en el período 1994-96, se ha reducido en 1998 hasta situarse en los niveles que tenía en 1.994. También en la Encuesta domiciliar sobre Drogas de 1.999 muestra que la proporción de españoles que habían consumido alguna vez éxtasis en el último año ha pasado a ser el 0,8% cuando en la anterior de 1.997 era del 1%. La importancia de este consumo ha provocado cierta alarma principalmente por la rapidez con que se ha producido su expansión y porque, aunque se trata de drogas cuyos efectos y toxicidad necesitan ser más investigados, existe suficiente evidencia

acerca de su problemática (Calafat, Sureda y Palmer, 1997; Calafat, Stocco, Mendes et al, 1998).

Con el presente estudio, se averiguará si las RNA pueden ser empleadas en un futuro como herramientas de apoyo al profesional dedicado a la prevención del consumo de este tipo de sustancias.

## MÉTODO

### Sujetos

La muestra estaba formada por dos grupos de sujetos, 148 consumidores de éxtasis y 148 no consumidores de éxtasis. El muestreo fue intencional, encuestándose a los jóvenes en los lugares recreativos donde acudían, y se realizó en cinco países de la Comunidad Europea: España, Francia, Holanda, Italia y Portugal. A su vez la muestra se podía dividir en función del lugar donde se había pasado el cuestionario: un grupo de usuarios de discoteca y otro de estudiantes de Universidad. En la tabla 1 se presentan las características demográficas de los sujetos consumidores y no consumidores.

El grupo de consumidores se caracterizaba por ser consumidores habituales de éxtasis —consumían éxtasis más de una vez al mes. En general, los sujetos que formaban esta categoría eran además consu-

**Tabla 1:**  
**Características demográficas de los sujetos consumidores y no consumidores de éxtasis.**

	<b>Consumidores</b> (n = 148)	<b>No consumidores</b> (n = 148)
<b>Sexo</b>		
<i>Mujer</i>	58	59
<i>Varón</i>	90	89
<b>Edad</b>	22.38* (4.15)	22.82* (4.30)
<b>País</b>		
<i>España</i>	48	34
<i>Francia</i>	21	29
<i>Holanda</i>	35	24
<i>Italia</i>	8	18
<i>Portugal</i>	36	43
<b>Lugar</b>		
<i>Discoteca</i>	108	69
<i>Universidad</i>	40	79

Nota: \* Media y desviación estándar.

midores de otras sustancias como marihuana (n = 118), cocaína (n = 70), anfetaminas (n = 51), LSD (n = 44) y heroína (n = 7). Por su parte, el grupo de no consumidores que ha servido como grupo control se caracterizaba por no haber consumido nunca éxtasis ni ninguna otra sustancia ilegal.

## Instrumentos

Con el objeto de determinar las características predictoras del consumo de éxtasis, se construyó un cuestionario compuesto por 25 ítems. Los ítems se podían agrupar en cinco categorías temáticas:

- a) Demografía, relaciones con los padres y creencias religiosas
- b) Ocio
- c) Consumo
- d) Opinión sobre el éxtasis
- e) Personalidad

Las áreas exploradas por este cuestionario coinciden en gran medida con los principios de la *Squashing Theory*, enfoque desarrollado por Buscema (1995) y encaminado a la predicción de la conducta adictiva, mediante un modelo de red neuronal, a partir del registro de un conjunto de medidas biológicas, psicológicas y sociológicas.

La naturaleza de los ítems del cuestionario era variada. La mayoría eran variables cualitativas polítmicas —p.e. “ocupación”—, pero había así mismo variables cualitativas dicotómicas —p.e. “¿eres creyente?”—, así como ítems de naturaleza ordinal —p.e. “estatus económico”—, e ítems de naturaleza cuantitativa —p.e. “puntuación en la escala de desviación social”.

El modelo de red neuronal utilizado en la parte empírica de este trabajo fue simulado en un ordenador PC mediante el programa Neural Connection 2.0 (SPSS Inc., 1997a), el cual permite implementar el algoritmo de aprendizaje *backpropagation* en una arquitectura del tipo perceptrón multicapa.

## Aplicación de la red neuronal

Resolver un problema mediante el uso de RNA supone aplicar una metodología que presenta aspectos comunes con las técnicas convencionales de modelado estadístico, pero también otros más particulares, que solamente se dan en el campo de las RNA. A continuación, se describen los pasos que se han seguido para la construcción de un modelo de red

neuronal capaz de discriminar entre sujetos consumidores o no consumidores de éxtasis.

### *Selección de las variables relevantes y preprocesamiento de los datos*

Para obtener una aproximación funcional óptima, se deben elegir cuidadosamente las variables a emplear. Más concretamente, de lo que se trata es de incluir en el modelo las variables predictoras que realmente predigan la variable dependiente, pero que a su vez no covaríen entre sí (Smith, 1993). La introducción de variables irrelevantes o que covaríen entre sí, puede provocar un sobreajuste innecesario en el modelo. Este fenómeno aparece cuando el número de parámetros o pesos de la red resulta excesivo en relación al problema a tratar y al número de patrones de entrenamiento disponibles. La consecuencia más directa del sobreajuste es una disminución sensible en la capacidad de generalización del modelo, es decir, la capacidad de la red de proporcionar una respuesta correcta ante patrones que no han sido empleados en su entrenamiento.

Teniendo en cuenta lo comentado, fue seleccionado un conjunto de 25 variables que permitían evaluar diferentes aspectos del sujeto, susceptibles de poder predecir el consumo de éxtasis. En la tabla 2 se proporciona una descripción de las variables predictoras utilizadas y la variable dependiente.

Una vez seleccionadas las variables que iban a formar parte del modelo, se procedió al preprocesamiento de los datos para adecuarlos a su tratamiento por la red neuronal. Para trabajar con el modelo de red neuronal aplicado en este estudio, el *backpropagation*, es muy aconsejable —aunque no imprescindible— conseguir que los datos posean una serie de cualidades (Masters, 1993; Martín del Brío y Sanz, 1997; SPSS Inc., 1997b; Sarle, 1998). Las variables deberían seguir una distribución normal o uniforme en tanto que el rango de posibles valores debería ser aproximadamente el mismo y acotado dentro del intervalo de trabajo de la función de activación empleada en las capas ocultas y de salida de la red neuronal.

Para adaptar nuestros datos a estas condiciones, se aplicó de forma satisfactoria una transformación logarítmica en las variables continuas que no seguían una distribución normal. A continuación, se acotó los valores de todas las variables predictoras al rango [-1, 1], límites de la función de activación que será utilizada por las neuronas de la capa oculta de la red. Este procedimiento permitió obtener mejores resultados que otros métodos de codificación comúnmente usados para el caso de variables cualitativas como, por ejemplo, los métodos 1-de-N y 1-de-N-1. Por su parte, la variable dependiente, estatus del sujeto, fue codificada como: -1 = no consumidor de éxtasis, 1 = consumidor de éxtasis.

**Tabla 2: Descripción de las variables predictoras y la variable dependiente.**

<b>Variable</b>	<b>Alternativas de respuesta</b>
<b>Variables predictoras</b>	
<i>Demografía, padres y religión</i>	
Estado civil	1: soltero/a 2: casado/a 3: vivo en pareja 4: otros
Nivel de estudios	1: primarios 2: bachiller 3: superiores
Ocupación	1: estudio 2: estudio y trabajo 3: trabajo eventual 4: trabajo fijo 5: servicio militar 6: parado 7: otros
Estatus económico	1: bajo 2: medio/bajo 3: medio 4: medio/alto 5: alto
¿Con quién vives?	1: padres/familia 2: conyuge/pareja 3: amigos 4: colegio/residencia 5: solo 6: otros
Relaciones con los padres	1: muy malas 2: bastante malas 3: regulares 4: bastante buenas 5: muy buenas
¿Eres creyente?	1: si 2: no
<i>Ocio</i>	
¿Vas a bares?	1: nunca 2: a veces 3: a menudo 4: casi siempre
¿Vas a discotecas?	1: nunca 2: a veces 3: a menudo 4: casi siempre
¿Vas a pubs?	1: nunca 2: a veces 3: a menudo 4: casi siempre
¿Vas a cafés?	1: nunca 2: a veces 3: a menudo 4: casi siempre
¿Vas a afters?	1: nunca 2: a veces 3: a menudo 4: casi siempre
¿Vas a fiestas raves?	1: nunca 2: a veces 3: a menudo 4: casi siempre
¿Qué tipo de música prefieres?	1: house-bacalao 2: hardcore 3: hardcore-house 4: mellow-house 5: rock 6: pop 7: otros
<i>Consumo</i>	
¿Cuántos amigos toman éxtasis?	1: ninguno 2: pocos 3: la mitad 4: casi todos 5: todos
¿Has consumido alcohol este último mes?	1: si 2: no
¿Has consumido tabaco este último mes?	1: si 2: no
¿Te has emborrachado este último mes?	1: no 2: una vez al mes 3: varias veces al mes 4: alguna vez por semana 5: una vez por semana 6: cada día
<i>Opinión sobre el éxtasis</i>	
¿Crees que el éxtasis puede crear problemas?	1: no 2: sí, es ilegal 3: sí, después mal 4: sí, crea adicción 5: sí, amigos no toman 6: sí, efectos imprevisibles 7: sí, adulteración 8: sí, problemas con familia 9: otros
¿Cuál crees que es la razón para consumir éxtasis?	1: relajarse 2: disfrutar de bailar 3: bailar más tiempo 4: estar mejor con otros 5: olvidar los problemas 6: sentirse bien 7: mejor sexo 8: estimular los sentidos
<i>Personalidad</i>	
Escala de emoción y búsqueda de aventuras	Puntuación entre 0 y 10
Escala de búsqueda de experiencias	Puntuación entre 0 y 10
Escala de desinhibición	Puntuación entre 0 y 10
Escala de susceptibilidad al aburrimiento	Puntuación entre 0 y 10
Escala de desviación social	Puntuación entre 0 y 10
<b>Variable dependiente</b>	
Estatus de consumo de éxtasis	1: consumidor (más de una vez al mes) 2: no consumidor

*Creación de los conjuntos de aprendizaje, validación y test*

En la metodología de las RNA, a fin de encontrar la red que tiene la mejor ejecución con casos nuevos — es decir, que sea capaz de generalizar—, la muestra de datos es a menudo subdividida en tres grupos (Bishop, 1995; Ripley, 1996): entrenamiento, validación y test.

Durante la etapa de aprendizaje de la red, los pesos son modificados de forma iterativa de acuerdo

con los valores del grupo de entrenamiento, con el objeto de minimizar el error cometido entre la salida obtenida por la red y la salida deseada por el usuario. Sin embargo, como ya se ha comentado, cuando el número de parámetros o pesos es excesivo en relación al problema —fenómeno del sobreajuste—, el modelo se ajusta demasiado a las particularidades irrelevantes presentes en los patrones de entrenamiento en vez de ajustarse a la función subyacente que relaciona entradas y salidas, perdiendo su habilidad de generalizar su aprendizaje a casos nuevos.

Para evitar el problema del sobreajuste, es aconsejable utilizar un segundo grupo de datos diferentes a los de entrenamiento, el grupo de validación, que permita controlar el proceso de aprendizaje. Durante el aprendizaje la red va modificando los pesos en función de los datos de entrenamiento y de forma alterada se va obteniendo el error que comete la red ante los datos de validación. De este modo, podemos averiguar cuál es el número de pesos óptimo, en función de la arquitectura que ha tenido la mejor ejecución con los datos de validación. Como se verá más adelante, mediante el grupo de validación también se puede determinar el valor de otros parámetros que intervienen en el aprendizaje de la red.

Por último, si se desea medir de una forma completamente objetiva la eficacia final del sistema construido, no deberíamos basarnos en el error que se comete ante los datos de validación, ya que de alguna forma, estos datos han participado en el proceso de entrenamiento. Se debería contar con un tercer grupo de datos independientes, el grupo de test el cuál proporcionará una estimación insesgada del error de generalización.

En el presente estudio, se obtuvieron estos tres conjuntos de datos mediante una asignación aleatoria de los 296 sujetos que formaban la muestra. Así, se contó con 176 sujetos de entrenamiento –de los cuales 88 eran consumidores y 88 eran no consumidores de éxtasis–, 60 sujetos de validación –de los cuales 30 eran consumidores y 30 eran no consumidores de éxtasis–, y 60 sujetos de test –de los cuales 30 eran consumidores y 30 eran no consumidores de éxtasis.

### Entrenamiento de la red neuronal

El modelo de red neuronal empleado ha sido una arquitectura del tipo perceptrón multicapa entrenada mediante la regla de aprendizaje *backpropagation* (propagación del error hacia atrás) (Rumelhart, Hinton y Williams, 1986). El perceptrón multicapa está formado por una capa de entrada, una capa de salida y una o más capas ocultas o intermedias; la información se transmite desde la capa de entrada hasta la capa de salida y cada neurona está conectada con todas las neuronas de la siguiente capa. La utilización del algoritmo *backpropagation* o alguna de sus múltiples variantes supone alrededor del 80% de las aplicaciones que se realizan con RNA (Caudill y Butler, 1992).

El funcionamiento de una red de este tipo consiste en el aprendizaje de un conjunto de pares de entradas y salidas de información dados como ejemplo, empleando un ciclo de propagación-adaptación compuesto por dos fases. En nuestro caso, la red debe aprender a relacionar los valores de las variables predictoras con el correspondiente estatus de consumo del sujeto. En la fase de propagación, se presenta a la capa de entrada de la red los valores de las 25 variables pre-

dictoras correspondientes a un sujeto de entrenamiento, esta información se va propagando a través de todas las capas superiores hasta generar una salida, se compara el resultado obtenido con la salida que se desea obtener — -1 si el sujeto es no consumidor y 1 si el sujeto es consumidor —, y se calcula el error que comete la neurona de la capa de salida. En la fase de adaptación, este error se propaga hacia atrás (de ahí el nombre que recibe), capa por capa, recibiendo cada neurona un error que describe su aportación relativa al error global que comete la red. Basándose en el valor del error recibido, se reajustan los pesos de conexión de cada neurona, de manera que en la siguiente vez que se presenten los valores del mismo sujeto, la salida esté más cerca de la deseada, es decir, el error disminuya.

A continuación, se expone la expresión matemática de la regla de modificación de pesos descrita (para una explicación más detallada, consultar: Rumelhart, Hinton y Williams, 1986):

$$\Delta w_{ji}(n+1) = \varepsilon \delta_{pj} x_{pi} + \eta \Delta w_{ji}(n)$$

donde

$w_{ji}$  = peso entre la neurona  $i$  y la neurona  $j$

$n$  = número de iteración

$\varepsilon$  = tasa de aprendizaje (junto al momento controla el tamaño del cambio de los pesos en cada iteración)

$\delta_{pj}$  = error de la neurona  $j$  para el patrón  $p$

$x_{pi}$  = salida de la neurona  $i$  para el patrón  $p$

$\eta$  = momento

Una vez que se han presentado todos los patrones de entrenamiento, se procede a actualizar el valor de los pesos de la red, completándose así un ciclo de aprendizaje o iteración. Con este proceso, se pretende minimizar la siguiente función de error:

$$E = \frac{1}{2} \sum_p \sum_k (d_{pk} - x_{pk})^2$$

donde

$d_{pk}$  = salida deseada de la neurona de salida  $k$  para el patrón  $p$

$x_{pk}$  = salida real de la neurona de salida  $k$  para el patrón  $p$

Es decir, el error que comete la red neuronal se obtiene calculando simplemente la diferencia entre la salida deseada por el usuario y la salida proporcionada por la red para cada patrón o sujeto de entrenamiento.

Antes de comenzar este proceso de aprendizaje, se debe asignar unos valores iniciales a los pesos de umbral y de conexión entre neuronas. Se adoptó el procedimiento común de asignar estos valores de

forma aleatoria dentro del rango [-0.5, 0.5] con una distribución uniforme (SPSS Inc., 1997a). Por otra parte, existe una serie de parámetros cuyo valor no se puede conocer *a priori* dado un problema, sino que deben ser determinados mediante ensayo y error. La utilización de un grupo de validación ayudará a conocer el valor óptimo de cada uno de estos parámetros: arquitectura de la red, valor de la tasa de aprendizaje y del momento, y función de activación de las neuronas de la capa oculta y de salida. Así, la configuración de parámetros que obtenga el menor error ante los datos de validación, será la seleccionada para pasar a la fase de test.

Respecto a la arquitectura de la red, se sabe que para la mayoría de problemas prácticos bastará con utilizar una capa de entrada, una oculta y una de salida (Funahashi, 1989; Hornik, Stinchcombe y White, 1989). El número de neuronas de la capa de entrada está determinado por el número de variables predictoras. Cada neurona de entrada tiene como misión recibir y transmitir a la siguiente capa, el valor de una de estas variables. Por su parte, el número de neuronas de la capa de salida está determinado, en tareas de clasificación, por el número de categorías o clases que tiene el problema. En nuestro caso, la única neurona de salida dará como resultado el valor -1 si el sujeto es no consumidor y 1 si el sujeto es consumidor. Por último, no existe una receta que indique el número óptimo de neuronas en la capa oculta para un problema dado. Recordando el problema del sobreajuste, se debe usar el mínimo número de neuronas ocultas con las cuales la red rinda de forma adecuada (Masters, 1993; Smith, 1993; Rzepoluck, 1998). Así, evaluando el rendimiento de diferentes arquitecturas en función de los resultados obtenidos con el grupo de validación, se seleccionó una capa oculta compuesta por dos neuronas.

Los valores de la tasa de aprendizaje ( $\epsilon$ ) y el momento ( $\eta$ ) tienen un papel crucial en el proceso de entrenamiento de una red neuronal, ya que controlan el tamaño del cambio de los pesos en cada iteración. Se deben evitar dos extremos: un ritmo de aprendizaje demasiado pequeño puede ocasionar una disminución importante en la velocidad de convergencia y la posibilidad de acabar con una configuración de pesos poco eficiente; en cambio, un ritmo de aprendizaje demasiado grande puede conducir a inestabilidades en la función de error o a saturar las neuronas de la red. Por tanto, se recomienda elegir un ritmo de aprendizaje lo más grande posible sin que provoque grandes oscilaciones. En general, el valor de la tasa de aprendizaje suele estar comprendida entre 0.05 y 0.5, mientras que el valor del momento suele ser aproximadamente igual a 0.9 (Rumelhart, Hinton y Williams, 1986). En nuestro estudio, los mejores resultados se obtuvieron con unos valores de  $\epsilon = 0.3$  y  $\eta = 0.8$ . Esta configuración de valores permitió alcanzar la conver-

gencia —es decir, hasta que el valor de los pesos permanece estable—, en 1200 iteraciones o ciclos de aprendizaje, momento en que se decidió parar el entrenamiento.

Por último, la función de activación es la función que se aplica a la entrada neta de la neurona para obtener un valor de salida. La entrada neta es la suma del producto de cada señal que recibe de las neuronas de la capa anterior por el valor del peso que conecta ambas neuronas, menos el umbral de la neurona (el umbral es considerado como un peso que conecta con una neurona ficticia con valor de salida igual a 1):

$$net_j = \sum_{i=1}^N w_{ji}x_i - \theta_j$$

El algoritmo *backpropagation* exige que la función de activación sea continua y, por tanto, derivable para poder obtener el error o valor delta de las neuronas ocultas y de salida. Se disponen de dos formas básicas que cumplen esta condición: la función lineal (o identidad) y la función sigmoideal. Sin embargo, es absolutamente imprescindible, para aprovechar la capacidad de las RNA de aprender relaciones complejas o no lineales entre variables, la utilización de funciones no lineales al menos en las neuronas de la capa oculta (Rzepoluck, 1998). En este sentido, los mejores resultados se obtuvieron utilizando la función sigmoideal tangente hiperbólica (con límites entre -1 y 1) para las neuronas de la capa oculta y la función lineal para la neurona de la capa de salida.

En consonancia con nuestros resultados, los estudios experimentales realizados muestran que la utilización de valores bipolares (positivos y negativos) en las funciones de activación acelera considerablemente el entrenamiento de la red frente a la utilización de valores binarios como es el caso de la función sigmoideal logística (con límites entre 0 y 1) (Fahlman, 1988; Kalman y Kwasny, 1992; Fausett, 1994).

La figura 1 muestra el modelo de red neuronal utilizado en este estudio, la arquitectura estaba compuesta por 25 neuronas en la capa de entrada, dos neuronas en la capa oculta y una neurona en la capa de salida.

#### *Evaluación del rendimiento de la red neuronal*

La evaluación del rendimiento de una RNA entrenada se realiza mediante el uso de un grupo de datos que no haya participado en el proceso de aprendizaje, el grupo de test. Con esto, se persigue obtener algún tipo de medida que permita estimar la capacidad de generalización del modelo. En este sentido, existe un amplio abanico de medidas de rendimiento (Masters, 1993): media cuadrática del error, funciones de coste, matrices de confusión, índices de sensibilidad y especificidad, etc..

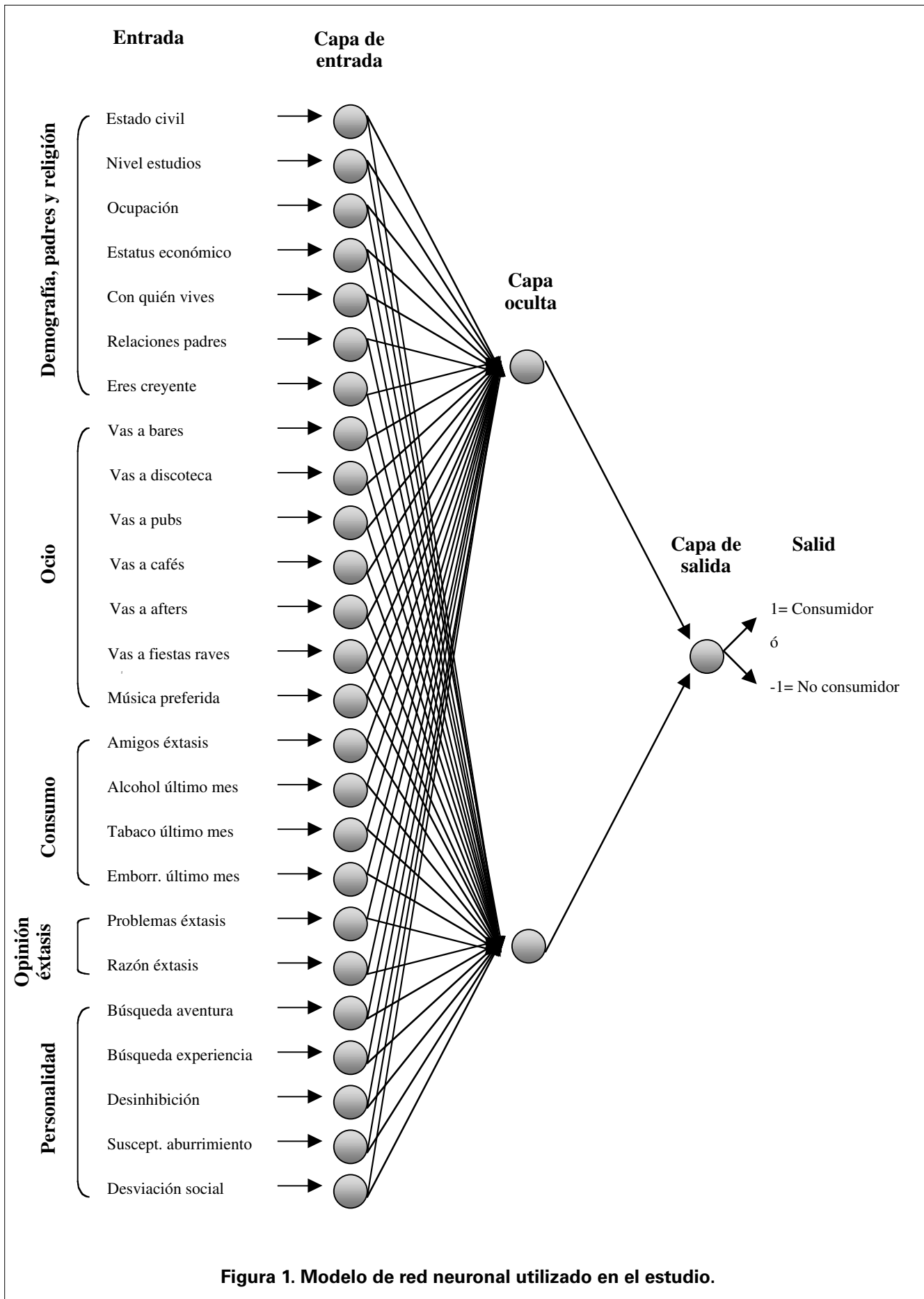


Figura 1. Modelo de red neuronal utilizado en el estudio.

En nuestro estudio, la evaluación del rendimiento se realizó a partir de los índices de sensibilidad, especificidad y eficacia, y del análisis de curvas ROC (*Receiver operating characteristic*).

Se recuerda al lector que la sensibilidad de un instrumento diagnóstico es, en nuestro caso, el porcentaje de consumidores que son clasificados correctamente —verdaderos positivos. Por su parte, la especificidad es el porcentaje de no consumidores que son clasificados correctamente —verdaderos negativos. Por último, a raíz de los dos índices anteriores, la eficacia es el porcentaje de sujetos (consumidores y no consumidores) correctamente clasificados.

El análisis de curvas ROC se originó a principios de los años 50 en el seno de la teoría de detección electrónica de señales (TDS), y se ha destacado en los últimos años como una medida precisa y válida para evaluar la precisión diagnóstica de un instrumento (Swets, 1973, 1988). Las curvas ROC poseen dos ventajas fundamentales respecto a los tradicionales índices de sensibilidad, especificidad y eficacia: son independientes del punto de corte elegido y de la prevalencia —en nuestro caso, de la proporción de sujetos consumidores. Para nuestros fines, la curva ROC consistiría en la representación gráfica del porcentaje de verdaderos positivos (sensibilidad) en el eje de ordenadas, contra el porcentaje de falsos positivos (1-especificidad) en el eje de abscisas, para diferentes puntos de corte aplicados sobre la salida que proporciona la red neuronal —un valor cuantitativo aproximadamente entre -1 y 1. Los verdaderos positivos serían sujetos consumidores clasificados por la red como consumidores, mientras que los falsos positivos serían sujetos no consumidores clasificados por la red como consumidores. En este tipo de análisis, la medida de resumen más utilizada es el área total bajo la curva ROC. Esta medida se interpreta como la probabilidad de clasificar correctamente un par de sujetos —uno consumidor y otro no consumidor—, seleccionados al azar, fluctuando su valor entre 0.5 y 1. El área bajo la curva ROC de un instrumento inútil es 0.5, reflejando que al ser utilizado clasificamos correctamente un 50% de individuos, idéntico porcentaje al obtenido utilizando simplemente el azar. Por el contrario, el área bajo la curva ROC de un instrumento perfecto es 1, ya que permite clasificar sin error el 100% de sujetos.

### *Análisis de sensibilidad*

Una de las críticas más importantes que se han lanzado contra el uso de RNA trata sobre lo difícil que es comprender la naturaleza de las representaciones internas generadas por la red para responder ante un problema determinado (De Laurentiis y Ravdin, 1994; Rzepoluck, 1998). A diferencia de los modelos estadísticos clásicos, no es tan evidente conocer en una

red la importancia que tiene cada variable predictora sobre la/s variable/s dependiente/s. Sin embargo, esta percepción acerca de las RNA como una compleja "caja negra," no es del todo cierta. De hecho, han surgido diferentes intentos por interpretar los pesos o parámetros del modelo (Masters, 1993), de los que el más ampliamente utilizado es el denominado análisis de sensibilidad (Hashem, 1992; Lisboa, Mehrdehnavi y Martin, 1994). Se debe advertir al lector que el término sensibilidad utilizado en el apartado anterior no tiene ningún tipo de relación con el término análisis de sensibilidad utilizado en esta ocasión. Recordemos que la sensibilidad es el porcentaje de verdaderos positivos de un instrumento diagnóstico, mientras que el análisis de sensibilidad es un procedimiento para conocer el efecto o influencia de cada variable predictora sobre la/s variable/s dependiente/s.

El método más común para realizar un análisis de sensibilidad consiste en fijar el valor de todas las variables de entrada a su valor medio e ir variando el valor de una de ellas a lo largo de todo su rango, con el objeto de observar el efecto que tiene sobre la salida de la red. Siguiendo este método, se fue registrando los cambios que se producían en la salida de la red cada vez que se aplicaba un pequeño incremento  $n$  —incrementos de un 2%—, en una variable de entrada. Se propuso como objetivo cuantificar la influencia que tiene cada variable de entrada. Pensamos que la simple suma de los cambios producidos proporcionaría una medida intuitiva de sensibilidad. Esta medida representaría el efecto relativo que tiene una variable de entrada sobre la salida de la red. Así, un valor cercano a 0 indicaría poco efecto o sensibilidad; a medida que se fuese alejando de 0, indicaría que el efecto va aumentando. Esta medida de sensibilidad se obtuvo mediante la siguiente expresión:

$$S_{ik} = \frac{N}{n} |x_{kn} - x_{kmin}|$$

donde

- $S_{ik}$  = medida de sensibilidad de la variable de entrada  $i$  sobre la salida  $k$
- $x_{kn}$  = valor de la salida  $k$  obtenido con el incremento  $n$  en la variable de entrada  $i$
- $x_{kmin}$  = valor de la salida  $k$  obtenido con el valor mínimo posible de la variable de entrada  $i$

## RESULTADOS

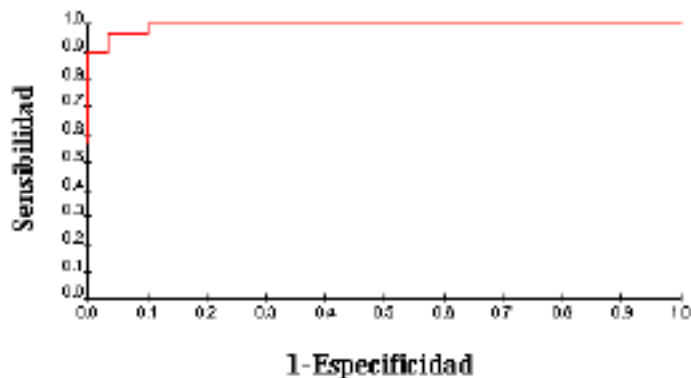
### Rendimiento del modelo de red neuronal

El modelo de red neuronal finalmente seleccionado obtuvo unos resultados excelentes a partir del



grupo de test –recordemos que este grupo estaba compuesto por 30 sujetos consumidores y 30 sujetos no consumidores de éxtasis. Así, estableciendo un punto de corte igual a cero en la salida de la red —las salidas negativas eran consideradas como “no consumidores” y las positivas como “consumidores”—, únicamente dos sujetos, uno de cada grupo, fueron

incorrectamente clasificados. Por tanto, los valores — en términos de porcentaje—, de la sensibilidad, especificidad y eficacia de la red fueron todos del 96.66%. Por su parte, el área total bajo la curva ROC (gráfico 1) dio como resultado  $0.9944 \pm 0.0055$  SE, aportando más datos a favor de la eficacia predictora del modelo entrenado.



**Gráfico 1. Curva ROC del modelo de red a partir del grupo de test.**

### Rendimiento de los submodelos de red neuronal

Una vez demostrado el excelente rendimiento del modelo de red entrenado, se quiso examinar la capacidad predictora de cada una de las cinco categorías temáticas —demografía, padres y religión, ocio, consumo, opinión sobre el éxtasis y personalidad—, sobre el consumo de éxtasis. Para ello, se crearon

cinco submodelos de red, cada uno entrenado a partir de las variables que formaban una categoría temática. Las condiciones de entrenamiento y evaluación fueron las mismas que las usadas para el modelo general de red utilizado inicialmente.

En la tabla 3 se presentan los índices de rendimiento de los cinco submodelos de red a partir del grupo de test.

**Tabla 3: Índices de rendimiento de los cinco submodelos de red a partir del grupo de test.**

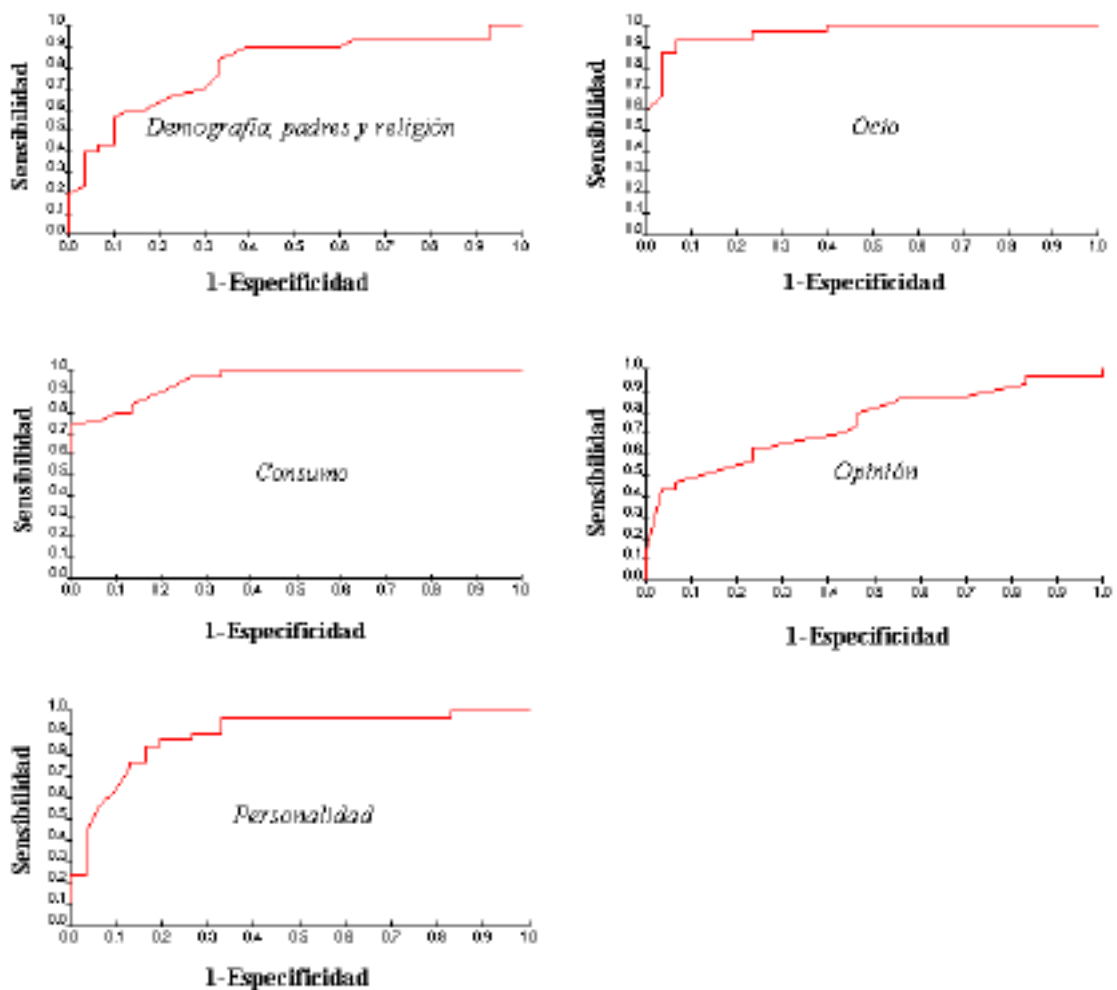
Categoría	Sensibilidad	Especificidad	Eficacia	Área ROC*
Demografía, padres y religión	80.00	66.66	73.33	0.80 (0.05)
Ocio	90.00	93.33	91.66	0.96 (0.02)
Consumo	90.00	80.00	85.00	0.95 (0.02)
Opinión sobre el éxtasis	46.66	93.33	70.00	0.74 (0.06)
Personalidad	90.00	70.00	80.00	0.88 (0.04)

Nota: \* Área bajo la curva ROC y error estándar.

Los valores de sensibilidad, especificidad, eficacia y área bajo la curva ROC indican que las dos categorías con mayor poder predictivo son las de ocio (91.66% de eficacia y 0.96 de área ROC) y consumo (85% de eficacia y 0.95 de área ROC). La categoría de personalidad alcanza un valor predictivo muy satisfactorio con una eficacia del 80% y un área ROC de 0.88. Por último, las categorías de demografía, padres y religión

(73.33% de eficacia y 0.80 de área ROC), y opinión sobre el éxtasis (70% de eficacia y 0.74 de área ROC) son las que presentan menor poder predictivo. Aunque la primera de ellas presenta una sensibilidad del 80% y la segunda presenta una especificidad del 93.33%.

En el gráfico 2 se muestra la curva ROC de cada uno de los cinco submodelos de red a partir del grupo de test.



**Gráfico 2: Curvas ROC de los cinco submodelos de red a partir del grupo de test.**

### Análisis de sensibilidad

A partir del modelo general inicialmente entrenado, se obtuvo el valor de la medida de sensibilidad para cada variable predictora sobre el consumo de éxtasis. En la tabla 4 se presentan estos valores ordenados de mayor a menor. Así, los primeros valores de la tabla corresponden a las variables de entrada con más influencia o relación con la salida de la red –estatus de consumo del sujeto.

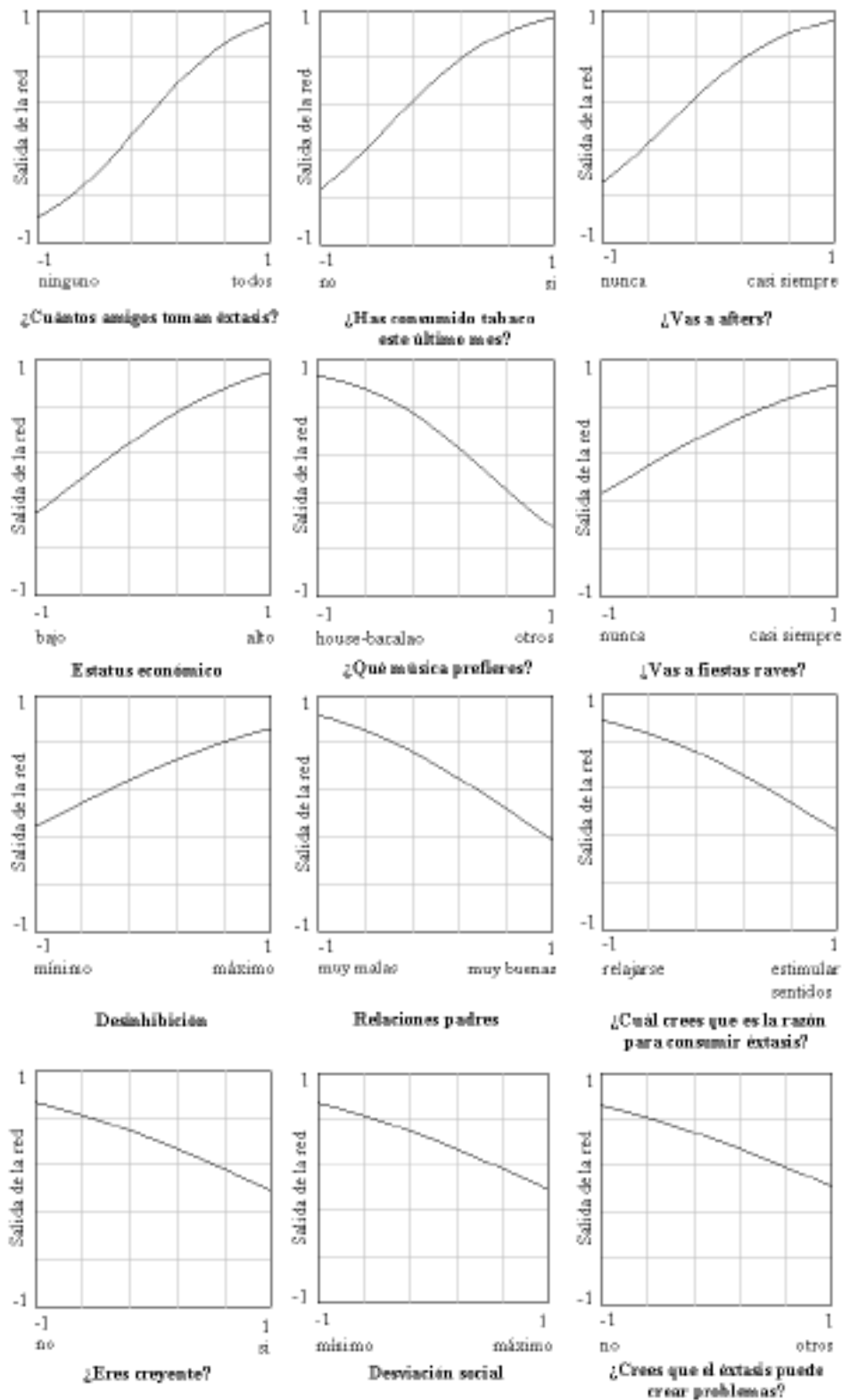
Así, se puede observar que las variables que tienen mayor influencia en el consumo de éxtasis son: la cantidad de amigos/as que consumen éxtasis ( $S = 58.93$ ), el consumo de tabaco ( $S = 43.18$ ), la frecuencia en asistir a afters ( $S = 41.24$ ), el estatus económico ( $S = 34.22$ ), el tipo de música preferida ( $S = 26.50$ ) y la frecuencia en asistir a fiestas raves ( $S = 26.21$ ). Estos resultados concuerdan con los obtenidos al evaluar el rendimiento de los diferentes submodelos, es decir, las variables de ocio y consumo son las que tienen mayor efecto sobre el consumo de éxtasis.

**Tabla 4: Medida de sensibilidad de las variables predictoras sobre el consumo de éxtasis.**

Variable predictora	Sensibilidad
¿Cuántos amigos toman éxtasis?	58.93
¿Has consumido tabaco este último mes?	43.18
¿Vas a afters?	41.24
Estatus económico	34.22
¿Qué tipo de música prefieres?	26.50
¿Vas a fiestas raves?	26.21
Escala de desinhibición	22.69
Relaciones con los padres	22.47
¿Cuál crees que es la razón para consumir éxtasis?	20.21
¿Eres creyente?	16.94
Escala de desviación social	15.90
¿Crees que el éxtasis puede crear problemas?	15.89
¿Has consumido alcohol este último mes?	12.31
Escala de susceptibilidad al aburrimiento	11.03
¿Vas a pubs?	10.19
Estado civil	9.42
Escala de emoción y búsqueda de aventuras	9.27
¿Vas a cafés?	7.34
Escala de búsqueda de experiencia	6.84
Ocupación	4.93
¿Con quién vives?	4.14
¿Vas a bares?	3.89
¿Tes has emborrachado este último mes?	2.80
Nivel de estudios	2.77
¿Vas a discotecas?	1.08

Por último, para obtener una información más completa, la medida de sensibilidad debería ir acompañada de la correspondiente representación gráfica. En el gráfico 3 se muestra la representación gráfica

del análisis de sensibilidad de las 12 primeras variables de la tabla 3, es decir, las 12 variables que muestran más influencia o relación con la salida de la red.



**Gráfico 3: Representación gráfica del análisis de sensibilidad de las 12 variables predictoras con mayor relación o influencia sobre el consumo de éxtasis.**

A modo de ejemplo, se puede observar en el citado gráfico el efecto o relación que mantiene la variable "¿Cuántos amigos toman éxtasis?" y la salida de la red —estatus de consumo del sujeto. Cuando dicha variable toma el valor  $-1$  (ningún amigo toma éxtasis), la salida de la red es aproximadamente igual a  $-1$  (no consumidor de éxtasis). A medida que se incrementa el valor de la variable de entrada ( $-0.5 =$  pocos,  $0 =$  la mitad,  $0.5 =$  casi todos consumen éxtasis), el valor de la salida de la red también va incrementándose. Finalmente, cuando la variable toma el valor  $1$  (todos mis amigos consumen éxtasis), la salida de la red es aproximadamente igual a  $1$  (consumidor de éxtasis). Por tanto, se puede decir que el número de amigos que consumen éxtasis está relacionado con la salida de la red y, por tanto, es un predictor del estatus de consumo del sujeto.

## CONCLUSIONES

Se ha presentado una RNA capaz de predecir el consumo de éxtasis a partir de las respuestas dadas a un cuestionario, con un grado de eficacia del 96.66%. Esto significa que conociendo las respuestas del sujeto a esas 25 preguntas, se puede anticipar si ese sujeto es consumidor o no de éxtasis, con un margen de error muy pequeño. Los resultados obtenidos, en nuestro estudio, son acordes con los obtenidos por el equipo de Buscema. Así, por ejemplo, Buscema, Intraligi y Bricolo (1998) desarrollaron varios modelos de red neuronal para la predicción de la adicción a la heroína. La eficacia de los diferentes modelos fue siempre superior al 91%, llegando a alcanzar, en algunos casos, el 97%. Por su parte, Maurelli y Di Giulio (1998) obtuvieron un modelo de red capaz de predecir el grado de alcoholismo de un sujeto, a partir de los resultados de varios tests biomédicos, con una capacidad de predicción del 93%. Todos estos resultados muestran que las excelentes cualidades exhibidas por las RNA en las diferentes disciplinas, se extienden al campo de las conductas adictivas.

Por otra parte, se ha pretendido mostrar, en contra de la concepción tradicional, que los pesos de un modelo de red neuronal pueden dar información acerca del grado de influencia de las variables de entrada sobre la salida de la red. De este modo, se ha mostrado que cuanto más alto sea el índice de sensibilidad ( $\lambda$ ) de una determinada variable de entrada, más relación o influencia ejercerá sobre la salida de la red —estatus de consumo o no consumo del sujeto. Intentos como el nuestro se encuentran en los trabajos de Modai, Saban, Stoler et al. (1995), los cuales identificaron mediante un análisis de sensibilidad los factores de buen pronóstico ante la aplicación de un tratamiento en pacientes psiquiátricos. Por su parte, Kashani, Nair,

Rao et al. (1996), con un esquema similar identificaron los factores asociados a las autoexpectativas negativas en adolescentes.

Por último, pensamos que los desarrollos futuros deberían ir encaminados hacia la aplicación de RNA en el resto de conductas relacionadas con el uso y abuso de sustancias —anfetaminas, cocaína, marihuana, etc.—, con el objeto de identificar los factores que influyen en cada una de estas conductas mediante el uso de índices de sensibilidad robustos. Los resultados de estos desarrollos podrían facilitar información importante a la hora de confeccionar programas de prevención de la conducta adictiva.

## REFERENCIAS BIBLIOGRÁFICAS

- Arbib, M.A. (Ed.) (1995). *The handbook of brain theory and neural networks*. Cambridge, Mass.: MIT Press.
- Arbib, M.A., Erdi, P. y Szentagothai, J. (1997). *Neural organization: structure, function and dynamics*. Cambridge, Mass.: MIT Press.
- Bishop, C.M. (1995). *Neural networks for pattern recognition*. New York: Oxford University Press.
- Buscema, M. (1995). Squashing Theory: A prediction approach for drug behavior. *Drugs and Society*, 8(3-4), 103-110.
- Buscema, M., Intraligi, M. y Bricolo, R. (1998). Artificial neural networks for drug vulnerability recognition and dynamic scenarios simulation. *Substance Use & Misuse*, 33(3), 587-623.
- Calafat, A., Sureda, M.P. y Palmer, A. (1997). Características del consumo de éxtasis en una muestra de universitarios y usuarios de discoteca. *Adicciones*, 9(4), 529-555.
- Calafat, A.; Stocco, P.; Mendes, et al (1998) *Characteristics and Social Representation of Ecstasy in Europe*. Palma de Mallorca. IREFREA.
- Caudill, M. y Butler, C. (1992). *Understanding neural networks: Computer explorations*. Cambridge, MA: MIT Press.
- De Laurentiis, M. y Ravdin, P.M. (1994). A technique for using neural network analysis to perform survival analysis of censored data. *Cancer Letters*, 77, 127-138.
- De Lillo, A. y Meraviglia, C. (1998). The role of social determinants on men's and women's mobility in Italy. A comparison of discriminant analysis and artificial neural networks. *Substance Use and Misuse*, 33(3), 751-764.
- Fahlman, S. (1988). *An empirical study of learning speed in back-propagation networks*. Tech. Rep. CMU-CS-88-162.
- Fausett, L. (1994). *Fundamentals of neural networks*. New Jersey: Prentice-Hall.
- Funahashi, K. (1989). On the approximate realization of continuous mapping by neural networks. *Neural Networks*, 2, 183-192.

- Hashem, S. (1992). Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions. *International Joint Conference on Neural Networks*, 419-424.
- Hilera, J.R. y Martínez, V.J. (1995). *Redes neuronales artificiales: Fundamentos, modelos y aplicaciones*. Madrid: Ra-Ma.
- Hornik, K., Stinchcombe, M. y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.
- Jang, J. (1998). Comparative analysis of statistical methods and neural networks for predicting life insurers' insolvency (bankruptcy) (The University of Texas at Austin, 1997). *Dissertation Abstracts International, DAI-A*, 59/01, 228.
- Kalman, B.L. y Kwasny, S.C. (1992). Why tanh? Choosing a sigmoidal function. *International Joint Conference on Neural Networks*, 578-581.
- Kashani, J.H., Nair, S.S., Rao, V.G., Nair, J. y Reid, J.C. (1996). Relationship of personality, environmental, and DICA variables to adolescent hopelessness: a neural network sensitivity approach. *Journal American Children and Adolescent Psychiatry*, 35(5), 640-645.
- Lisboa, P., Mehridehnavi, A. y Martin, P. (1994). The interpretation of supervised neural networks. *Proceedings of the Workshop on Neural Network Applications and Tools*, 11-17.
- Martín del Brío, B. y Sanz, A. (1997). *Redes neuronales y sistemas borrosos*. Madrid: Ra-Ma.
- Massini, G. y Shabtay, L. (1998). Use of a constraint satisfaction network model for the evaluation of the methadone treatments of drug addicts. *Substance Use & Misuse*, 33(3), 625-656.
- Masters, T. (1993). *Practical neural networks recipes in C++*. London: Academic Press.
- Maurelli, G. y Di Giulio, M. (1998). Artificial neural networks for the identification of the differences between "light" and "heavy" alcoholics, starting from five nonlinear biological variables. *Substance Use & Misuse*, 33(3), 693-708.
- Modai, I., Saban, N.I., Stoler, M., Valevski, A. y Saban, N. (1995). Sensitivity profile of 41 psychiatric parameters determined by neural network in relation to 8-week outcome. *Computers in Human Behavior*, 11(2), 181-190.
- Palmer, A. y Montaña, J.J. (1999). ¿Qué son las redes neuronales artificiales?. Aplicaciones realizadas en el ámbito de las adicciones. *Adicciones*, 11(3), 243-255.
- Plan Nacional sobre Drogas (2.000). *Informe nº 3. Observatorio español sobre drogas*. Madrid: Plan Nacional sobre Drogas.
- Ripley, B.D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.
- Rumelhart, D.E., Hinton, G.E. y Williams, R.J. (1986). Learning internal representations by error propagation. En: D.E. Rumelhart y J.L. McClelland (Eds.). *Parallel distributed processing* (pp. 318-362). Cambridge, MA: MIT Press.
- Rzempoluck, E.J. (1998). *Neural network data analysis using Simulnet*. New York: Springer-Verlag.
- Sarle, W.S. (Ed.) (1998). *Neural network FAQ*. Periodic posting to the Usenet newsgroup comp.ai.neural-nets, URL: <ftp://ftp.sas.com/pub/neural/FAQ.html>.
- Simpson, P.K. (Ed.) (1995). *Neural networks technology and applications: theory, technology and implementations*. New York: IEEE.
- Smith, M. (1993). *Neural networks for statistical modeling*. New York: Van Nostrand Reinhold.
- Speri, L., Schilirò, G., Bezzetto, A., Cifelli, G., De Battisti, L., Marchi, S., Modenese, M., Varalta, F. y Consigliere, F. (1998). The use of artificial neural networks methodology in the assessment of "vulnerability" to heroin use among army corps soldiers: A preliminary study of 170 cases inside the Military Hospital of Legal Medicine of Verona. *Substance Use & Misuse*, 33(3), 555-586.
- SPSS Inc. (1997a). *Neural Connection 2.0* [Programa para ordenador]. SPSS Inc. (Productor). Chicago: SPSS Inc. (Distribuidor).
- SPSS Inc. (1997b). *Neural Connection 2.0: User's Guide* [Manual de programa para ordenadores]. Chicago: SPSS Inc.
- Swets, J.A. (1973). The relative operating characteristic in psychology. *Science*, 182, 990-1000.
- Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- Waller, N.G., Kaiser, H.A., Illian, J.B. y Manry, M. (1998). A comparison of the classification capabilities of the 1-dimensional Kohonen neural network with two partitioning and three hierarchical cluster analysis algorithms. *Psychometrika*, 63(1), 5-22.
- West, P., Brockett, P. y Golden, L. (1997). A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, 16(4), 370-391.

