

ing is carried out dynamically according to the experience acquired during the classification of the system. Validation is applied using 12 case bases and 6 individual classifiers and compared with the most known models in the literature. Statistical tests were used to determine that the proposed modification is significantly superior to the original one and it was even in some cases the results were superior to those obtained with Stacking.

KeyWords: multiple classifier system, majority vote, classification experience.

1. INTRODUCCIÓN

Los problemas de clasificación han cobrado particular importancia en la actualidad puesto que en muchos casos se precisa de agrupar elementos que presenten un conjunto de características específicas. Existen varios modelos reportados en la literatura como redes bayesianas, redes neuronales, árboles de decisión y modelos estadísticos, por solo mencionar algunos. Todos ellos con sus particularidades en la forma de realizar el aprendizaje para poder clasificar la mayor cantidad de ejemplos posibles de forma correcta.

En determinadas áreas de investigación, como la bioinformática y quimioinformática, la utilización de sistemas multiclasificadores en la solución de problemas de reconocimiento de patrones y de clasificación en sentido general, está siendo usada con bastante frecuencia ya que es preferible combinar la salida de varios clasificadores a utilizarlas de forma individual [1]. La idea de combinar varios clasificadores está justificada por el hecho de que combinar el criterio de varios expertos puede conducir a obtener un resultado que puede no ser el mejor, pero al menos evita tomar la peor decisión. Además, es posible que los expertos no tomen la decisión óptima de forma aislada pero la combinación de sus criterios puede dar como consecuencia una decisión no existente inicialmente y que puede aproximarse a la óptima [2].

Existen varias formas en las que se puede construir un sistema multiclasificador, pero todas ellas coinciden en dos elementos fundamentales; la selección de los clasificadores base y el modo en que se van a combinar sus salidas [3]. Entre los modelos más populares que combinan clasificadores se encuentran *Stacking*, *Bagging* y *Boosting*.

Teóricamente, una de las condiciones para obtener buenos resultados con un multiclasificador es garantizar la diversidad entre los clasificadores base. Lo anterior está justificado por la idea de que si todos los clasificadores asignaran la misma clase a las instancias, no tendría sentido combinarlos, ya que el objetivo es lograr una clasificación correcta en aquellas instancias donde un clasificador individual falla. Esta diversidad puede ser generada de diferentes formas, ya sea en la selección de los

clasificadores de base, en la selección de rasgos que serán considerados en la clasificación o en la formación de distintos conjuntos de entrenamiento [4].

Uno de los enfoques más simples considerados en la construcción de un sistema multiclasificador es el método del voto de clasificadores donde, a partir de la clasificación individual de un número de clasificadores, se decide la clase final de una instancia según algún criterio que puede ser el promedio de las probabilidades de cada clase, el máximo, el mínimo, y el producto de las mismas o por voto mayoritario [5]. Adicionalmente, esta forma de voto puede ser modificada para que cada clasificador se especialice en un conjunto de rasgos [6], o para que en el momento de la votación el criterio de los clasificadores sea considerado o no de acuerdo a un peso establecido por algún criterio. En [1], Kuncheva propone varias formas de lograr un voto pesado entre los clasificadores individuales a partir del desempeño individual que tienen durante la clasificación o considerando de alguna forma la relación del mejor clasificador con el resto de los clasificadores individuales. Sin embargo, los modelos propuestos en este caso no consideran el conocimiento adquirido por cada clasificador durante el proceso de entrenamiento e incluso en la posterior clasificación.

El presente trabajo se centra en desarrollar una modificación al voto mayoritario tradicional para asignarle pesos a cada clasificador individual teniendo en cuenta la experiencia adquirida durante la clasificación del sistema y haciendo que la forma de votación sea dinámica. Todo esto sustentado en la idea de que el comportamiento del sistema pueda adaptarse según el comportamiento de sus miembros.

2. MATERIALES Y MÉTODOS

Dado que el voto mayoritario es una de las 6 reglas de combinación implementadas en el multiclasificador Vote existente en el WEKA [1], se diseña un experimento que compare la modificación propuesta con el voto mayoritario implementado y con el mejor clasificador individual de la combinación formada en el multiclasificador. Adicionalmente, se

estudia el desempeño del voto dinámico dentro del multclasificador *Vote* con respecto a otros multclasificadores más populares.

2.1 Bases de casos

Para validar la modificación que se propone al voto mayoritario, se seleccionaron 12 bases de casos del repositorio de aprendizaje automático de la Universidad de California Irvine UCIML (UCI Repository of Machine-Learning Databases) [7]. Las bases usadas son *Echocardiogram*, *Heart statlog*, *Breast cáncer Wisconsin*, *House votes*, *Hepatitis*, *Tic Tac Toe*, *Australian*, *Diabetes*, *German credit*, *Pro Ortolology*, *Ionosphere*, y *Sonar*.

- *Australian* es una base con 690 casos, definidos por 14 rasgos y distribuidos en dos clases
- *Breast cáncer Wisconsin* recoge ejemplos de casos con cáncer de pecho. En esta base se describen las células estudiadas según 9 rasgos y se agrupan en 2 clases
- *Diabetes* está descrita por 8 rasgos continuos y recopila 768 casos con descripciones de posibles enfermos de diabetes mellitus
- *Echocardiogram* es una base definida por 12 rasgos distribuidos en dos clases y está conformada a partir del estudio realizado en 132 pacientes para determinar si sobreviven al menos un año después de un ataque al corazón
- *German credit* contiene 1000 casos, descritos por 13 rasgos nominales y distribuidos en dos clases
- *Heart statlog* es una base de caso que representa la presencia o no de enfermedades del corazón y cuenta con 270 instancias distribuidas en dos clases.
- *Hepatitis* es una base que cuenta con 155 casos que describen las características de pacientes que tienen esta enfermedad y los clasifica según su supervivencia
- *House votes* es una base que contiene 435 casos, con 15 rasgos nominales agrupados en dos clases
- *Pro Ortolology* es una base formada a partir de 4314 con 11 rasgos numéricos y divididos en dos clases
- *Tic Tac Toe* es una base formada a partir de movimientos realizados en el juego de mismo nombre, con 958 casos y agrupados en las categorías de ganar y perder
- *Ionosphere* es una base formada a partir de 34 tipos de mediciones efectuadas en 351 análisis de retorno de señal de radar desde la ionósfera
- *Sonar* contiene 208 experimentos realizados

para clasificar señales de sonar según el objeto contra el que rebotan. La clasificación está dada por las categorías de 'roca' o 'metal'

2.2 Modificación del voto mayoritario para considerar la experiencia de los clasificadores dentro de un sistema multclasificador

El multclasificador *Vote*, implementado en el WEKA con el mismo nombre, utiliza para generar la clasificación de una instancia el criterio de varios expertos (clasificadores) [1]. La salida de cada clasificador individual está dada por vectores con una distribución de probabilidad para cada una de las clases del problema. La combinación de estas probabilidades utiliza diferentes criterios como el promedio, mínimo, máximo, producto o mediana de las probabilidades. Otro criterio puede estar dado por el voto dado por los expertos en cada clase, tomándose aquella que acumule la mayor cantidad de votos.

Para este último caso, considérese L clasificadores $(C_1, C_2, C_3, \dots, C_L)$ para un problema de clasificación de m clases $(w_1, w_2, w_3, \dots, w_m)$. Un clasificador C_j tiene para la clase w_i un estimado de probabilidad $P(C_j, w_i), i = 1, 2, \dots, m$. De esto último se tiene que:

$$\sum_{i=1}^m P(C_j, w_i) = 1 \quad (1)$$

El estimado de probabilidad de un conjunto de clasificadores para cada una de las clases de un caso X se puede representar por la matriz MP , como se muestra en (2).

$$MP = \begin{bmatrix} P(C_1, w_1) & \dots & P(C_1, w_i) & \dots & P(C_1, w_m) \\ \dots & \dots & \dots & \dots & \dots \\ P(C_j, w_1) & \dots & P(C_j, w_i) & \dots & P(C_j, w_m) \\ \dots & \dots & \dots & \dots & \dots \\ P(C_L, w_1) & \dots & P(C_L, w_i) & \dots & P(C_L, w_m) \end{bmatrix} \quad (2)$$

Donde las columnas representan el estimado de cada clasificador para la i -ésima clase y las filas están asociadas a la salida dada por el j -ésimo clasificador. Según la matriz descrita en (2), se define el voto mayoritario a partir de la formación de una matriz MK donde cada posición se calcula como:

$$MK_{j,i} = \begin{cases} 1 & \text{si } MP(j, i) = \max_{1 \leq i \leq m} MP(j, i) \\ 0 & \text{en otro caso} \end{cases} \quad (3)$$

El voto para una clase estaría dado por (4) y la clase tomada utilizando voto mayoritario sería calculada según (5).

$$Voto_{w_i} = \sum_{j=1}^L MK(j, i) \quad (4)$$

$$Clase_{VM} = w_i, \text{ donde } i = \max_{1 \leq i \leq m} \sum_{j=1}^L MK(j, i) \quad (5)$$

En caso de la existencia de empates entre los votos dados por los expertos, se selecciona aleatoriamente la decisión a tomar.

En este método, solamente se analizan los criterios dados por cada clasificador en la clasificación de la instancia con la que están trabajando y no tienen en cuenta la experiencia o el historial de clasificación de cada modelo individual. Por ejemplo, el hecho de que un clasificador tienda a equivocarse durante la clasificación de $i-1$ instancias, al clasificar correctamente la instancia i puede no considerarse su criterio dado que es un comportamiento que históricamente no es común en él. Dicho modo de actuar puede encontrarse comúnmente en la forma de pensar de las personas ya que por lo general, existe cierta desconfianza ante lo que opina una persona que normalmente no piensa como lo que está diciendo.

Analizando lo anterior, se puede introducir la idea de un multclasificador controlado por una entidad que se encargue de decidir qué opinión se debe tener en cuenta para dar la clasificación final. En la Figura. 1 se puede observar un esquema general del funcionamiento de este multclasificador.

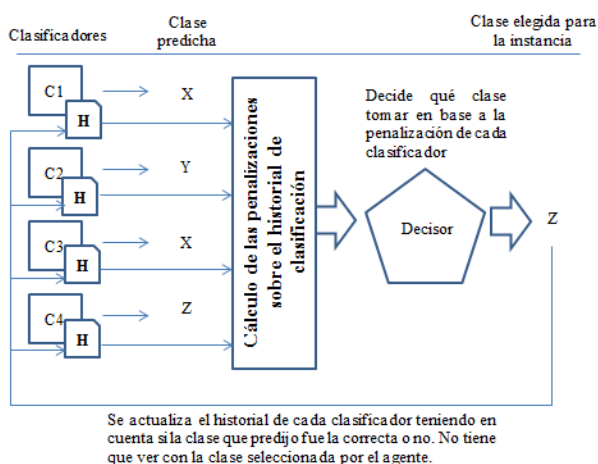


Figura. 1: Esquema general del multclasificador basado en la experiencia de los clasificadores individuales

En este tipo de multclasificador, el elemento esencial para garantizar un adecuado funcionamiento radica en la elección de la penalización que se define para la consideración o no de la clase predicha por un clasificador y la forma en que el *decisor* asigna la clase final a la instancia. La variante de penalización que se propone en este trabajo es la siguiente:

$$P_{C_j} = \frac{\sum_{i=1}^n I_i}{n},$$

$$I_i = \begin{cases} 1 & \text{clasificación incorrecta de la instancia} \\ 0 & \text{en otro caso} \end{cases} \quad (6)$$

donde:

n : total de instancias clasificadas hasta el momento

Esta penalización se basa sobre la razón que representa la cantidad de instancias incorrectamente clasificadas de las instancias clasificadas hasta el momento. Por tanto, la decisión a tomar sería considerar el criterio de aquellos clasificadores que menos penalización tengan. Esta penalización es similar a la propuesta por Kuncheva en [1], pero en vez de utilizarla solamente al final del entrenamiento de los clasificadores, en la variante propuesta se varía su valor según las instancias que se vayan evaluando, dando la posibilidad de que los clasificadores que tendían a no ser considerados por tener un mal desempeño, sí sean tenidos en cuenta posteriormente si su desempeño fue mejorando. La diferencia radica, fundamentalmente, en la forma en que es construida la penalización y cómo afecta el resultado final del sistema multclasificador.

Adicional a esta penalización pudiese agregarse un factor de optimismo de forma que le pueda dar a un clasificador que se ha equivocado varias veces la oportunidad de considerarse un poco más su criterio a pesar de que su penalización diga lo contrario. Este factor de optimismo estaría entre 0 para considerar la penalización nula y por tanto el criterio del clasificador tendría más oportunidad de ser considerado en la decisión final (se convertiría en un voto mayoritario), y un factor de 1 para considerar la penalización tal y como fue calculada. Este factor de optimismo sería agregado a la forma en que el decisor asigna la clase final a la instancia. En este trabajo es considerado 1 como factor de optimismo para tener en cuenta exactamente la penalización como fue calculada.

Finalmente, la fórmula observada en (3), debe ser modificada para considerar la penalización propuesta de la siguiente forma

$$MK_{j,i} = \begin{cases} 1 - \frac{\sum_{i=1}^n I_i}{n}, & \text{si } MP(j,i) = \max_{1 \leq i \leq m} MP(j,i) \\ 0 & \text{en otro caso} \end{cases} \quad (7)$$

2.2.1 Selección de los clasificadores base

La selección de los modelos de clasificación que serán usados en la construcción del multiclasificador, es una tarea compleja dada la gran variedad de clasificadores existente en la literatura. Una estrategia estándar puede ser seleccionar el mejor clasificador y guiarse por las decisiones que tome este o utilizar algún estimador que permita determinar dinámicamente un clasificador de acuerdo a su vecindad [1]. Muchos autores han propuesto realizar la selección de los clasificadores de forma que aseguren una diversidad en la clasificación individual y que pueda ser aprovechada en la clasificación final del multiclasificador [2, 8-10]. En muchas investigaciones se utilizan meta heurísticas para buscar un subconjunto de modelos que aseguren una mayor exactitud que la que pueda obtenerse de forma individual [11-14]. No obstante, el presente trabajo no pretende optimizar la búsqueda de los clasificadores, solo se propone una nueva forma de combinar las salidas de un conjunto con el objetivo de buscar una mejor región de decisión.

Atendiendo a lo anterior, la selección de los clasificadores fue efectuada de acuerdo a diferentes tipos de métodos de clasificación. Fue seleccionado un clasificador bayesiano (*Naive Bayes*, *NB*), un modelo de regresión (*Logistic*, *L*), un modelo perezoso (*kNN* con $k=3$, *IBk3*), una red neuronal (*Multi-Layer Perceptron*, *MLP*), y dos árboles de decisión; *ADTree* (*ADT*) y *Random Forest* (*RF*). Este último también puede ser considerado como una especie de multiclasificador según el funcionamiento propuesto en *Bagging* [15, 16].

Cada uno de estos clasificadores fue configurado con las opciones por defecto que establece WEKA, construyéndose un total de 57 multiclasificadores a partir de todas las combinaciones posibles (se omiten aquellas en las que solamente hay un clasificador) entre 6 clasificadores y aplicadas en las 12 bases de estudio.

2.3 Comparación con otros sistemas multiclasificadores

Además de efectuar la validación de la modificación propuesta con el mejor clasificador individual del multiclasificador y con la regla de voto mayoritario implementada en el Vote incluido en el WEKA, se decide analizar el comportamiento de los resultados obtenidos con respecto a otros sistemas multiclasificadores más populares como *Bagging*, *Boosting* y

Stacking.

Bagging fue introducido por Breiman en 1996 [15] y su esencia radica en entrenar un único clasificador en diferentes réplicas bootstrap¹. La clasificación dada será aquella que más votos tenga dentro del conjunto total de clasificadores (considérese cada entrenamiento en las réplicas como un clasificador aunque se utilice un mismo modelo). Para aprovechar las variaciones presentes en el conjunto de entrenamiento es necesario que el clasificador base que se utilice sea inestable; de otra forma el sistema construido sería una colección de clasificadores casi idénticos y sería muy improbable mejorar el desempeño que pueda tener uno solo de ellos [1].

Combinar múltiples modelos solo es válido si son significativamente diferentes y cuando cada uno de ellos maneja los datos de forma razonablemente correcta. Idealmente, los modelos deben complementarse de forma que cada uno de ellos pueda decidir qué hacer cuando el comportamiento de los otros no sea muy bueno [17]. *Boosting* explota esta idea buscando modelos complementarios usando el voto para combinar la salida de modelos individuales de un mismo tipo [4]. La diferencia con respecto a *Bagging* radica en que cada modelo formado es influenciado por el desempeño de los construidos previamente; midiendo, en forma de pesos, la contribución de cada modelo de forma que cada uno de ellos sea un experto en las instancias manejadas incorrectamente por los otros.

El otro modelo de combinación de clasificadores contra el cual se valida la modificación del voto mayoritario propuesto, es *stacked generalization*, o simplemente *Stacking*. A diferencia de los dos vistos anteriormente, *Stacking* combina modelos de clasificación no necesariamente del mismo tipo y para ello utiliza el concepto de *meta-learner*, que reemplaza el procedimiento del voto visto en *Bagging* y *Boosting*. El éxito de *Stacking* radica en su habilidad de explotar la diversidad en las predicciones de los clasificadores individuales que utiliza [5].

Para efectuar la validación, se obtuvieron los resultados de aplicar *Bagging* y *Boosting* tomando cada uno de los clasificadores individuales definidos anteriormente y seleccionando el de mejor desempeño de acuerdo a cada combinación formada. Por ejemplo, en la combinación NB-L-MLP, si la configuración de *Bagging* usando la red neuronal fue la mejor de las otras dos, se selecciona este para comparar con *Stacking* y *Vote* formado a partir de

¹ El bootstrapping es un método de remuestreo que en *Bagging* es usado para generar diferentes conjuntos de entrenamiento de un mismo tamaño a partir de uno original, mediante la eliminación de instancias y replicación de otras

esa combinación. Otro aspecto a tener en cuenta es la selección del *meta-learner* en *Stacking*, que en este caso fue seleccionada una red neuronal (*Multi-Layer Perceptron*), por su potencia en problemas de clasificación [3].

Resumiendo, la validación queda definida a partir de la formación de 6 grupos experimentales; uno que contiene los mejores clasificadores individuales en cada combinación, otro para los multclasificadores combinados según voto mayoritario, uno para la configuración de *Stacking* formada, otros dos para las configuraciones de *Bagging* y *Boosting* formadas según lo explicado anteriormente y un último grupo con los clasificadores combinados según el voto dinámico.

2.3.1 Evaluación de la clasificación

Las medidas más conocidas para evaluar la clasificación están basadas en la matriz de confusión que se obtiene cuando se prueba el clasificador en un conjunto de datos que no intervienen en el entrenamiento. Algunas de ellas pueden ser la exactitud (o *accuracy*), la sensibilidad, la precisión, la especificidad o el error. Otras formas de evaluar el desempeño de los clasificadores es por el análisis de las llamadas curvas ROC formadas a partir de la representación gráfica de la sensibilidad frente a la especificidad o calculando el área bajo esta curva (AUC).

Otro aspecto a tener en cuenta es la forma en que son entrenados los modelos de clasificación. Dentro de los existentes en la literatura [18], el método de validación cruzada con k subconjuntos (*k-fold cross-validation*) es uno de los más usados. Este método se basa en dividir la base de casos en k segmentos y realizar k procesos de entrenamientos y pruebas, de forma que en cada proceso se tome un subconjunto diferente de los formados para prueba y el resto para entrenar. Dado que el voto dinámico propuesto es sensible al orden en que son presentados los casos para aprender, fue utilizado este método para realizar la validación tomando $k=10$ según lo expresado en [19] y la exactitud de los modelos de clasificación como medida de evaluación.

La propuesta de modificación al voto mayoritario será denominada en lo sucesivo como VD (voto dinámico).

3. RESULTADOS Y DISCUSIÓN

En la Tabla I se muestra la exactitud de los clasificadores individuales usados en la experimentación, resaltándose en negritas aquellos que mejores resultados obtuvieron. Puede observarse que aunque no sobresalió ninguno si se obtuvieron buenos resultados en el clasificador bayesiano, el de regresión logística y el *Random Forest*.

Tabla I: Exactitud de los clasificadores individuales usados en la experimentación (%)

	Naive Bayes (NB)	Logistic (L)	kNN-3 (IBk3)	Random Forest (RF)	Multi-Layer Perceptron (MLP)	ADTree (ADT)
Australian	77.6812	86.087	84.2029	87.971	83.7681	85.5072
Breast-cancer-wisconsin	97.3646	93.9971	95.3148	96.3397	95.9004	95.3148
Diabetes	76.3021	77.2135	72.6563	73.4375	75.3906	72.9167
Echocardiogram	68.1818	72.7273	65.9091	69.697	68.1818	64.3939
German credit	75.4	75.2	73.3	74	71.5	72.4
Heart-statlog	83.7037	83.7037	78.8889	81.1111	78.1481	78.5185
Hepatitis	84.5161	82.5806	81.2903	85.1613	80	76.129
House votes	90.1149	96.092	92.6437	96.7816	95.1724	96.3218
Pro Ortology	80.7835	82.1511	80.4358	79.8331	82.128	82.3134
Tic Tac Toe	69.6242	98.3299	98.7474	91.1273	96.8685	76.096
Ionosphere	82.6211	88.8889	86.6097	94.5869	91.1681	93.1624
Sonar	67.7885	73.0769	86.0577	80.2884	82.2115	77.8846

En la Figura. 2 se muestran gráficamente los resultados obtenidos utilizando voto dinámico en los 57 multclasificadores formados a partir de las combinaciones de los clasificadores individuales (aplicados en las 12 bases de estudio) y comparados con, de izquierda a derecha, el mejor individual de la combinación (MI), voto mayoritario (VM), *Stacking*

(S), *Bagging* (Ba) y *Boosting* (Bo).

Cada punto representa una combinación de clasificadores (de las 57 posibles) aplicada en una de las 12 bases de estudio para formar una nube de 684 puntos. Por ejemplo, en la primera gráfica se muestra el valor de la exactitud del mejor clasificador individual según la combinación formada por el eje

de las "x" y el valor de la exactitud de VD por las "y". Un punto encima de la recta $y=x$ significa que el resultado obtenido combinando las salidas con voto dinámico es superior y un punto por debajo de la recta significa lo contrario. Aunque los resultados no son apreciables a simple vista, se puede observar que con respecto a voto mayoritario, la modifi-

cación propuesta a este es superior (y muy similares) en muchos casos y que los resultados obtenidos en *Stacking* están por debajo de los obtenidos por voto dinámico en aproximadamente más del 50% de los casos.

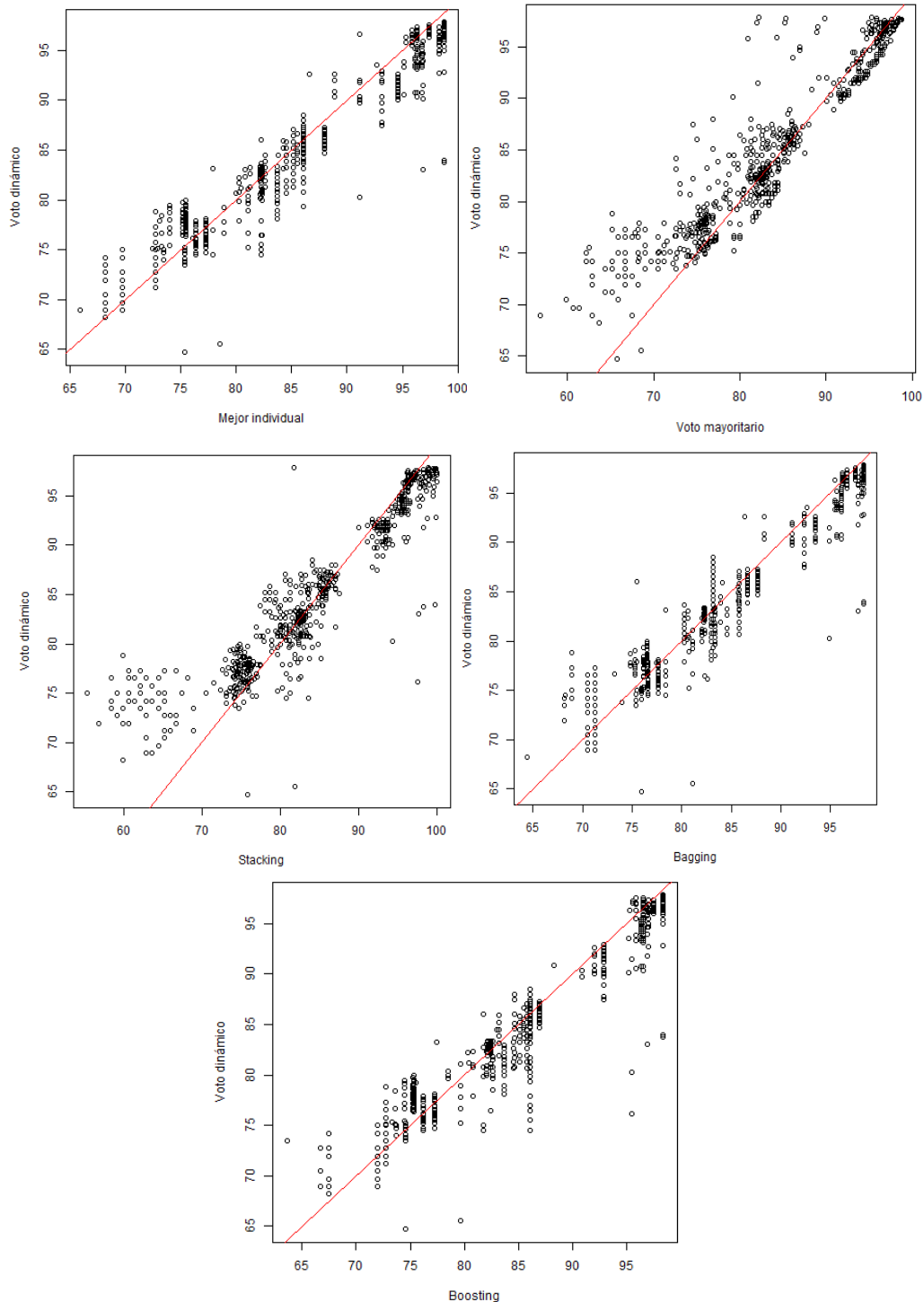


Figura. 2: Comparación de VD contra MI, VM, S, Ba y Bo

Sin embargo, los resultados obtenidos en *Bagging* y *Boosting* son superiores a los obtenidos con la modificación propuesta. En el caso particular de *Bagging*, los resultados obtenidos pueden deberse a las características propias de este modelo ya que para aprovechar las variaciones presentes en lo conjuntos de entrenamiento, es necesario que el clasificador que se utilice sea inestable para que pequeñas modificaciones en el conjunto de entrenamiento puedan conducir a grandes cambios en la salida del clasificador. Al analizarse la Tabla II, puede observarse que los mejores resultados estuvieron en aquellas configuraciones en que fueron usados árboles de decisión y la red neuronal (ambos modelos considerados inestables).

Tabla II: Exactitud de *Bagging* tomando cada uno de los clasificadores individuales (%)

Base	NB	L	IBk3	RF	MLP	ADT
1	77.39	86.67	84.06	87.68	85.80	85.65
2	97.51	96.19	95.31	96.19	96.63	95.46
3	76.56	77.60	73.96	75.39	65.10	75.91
4	68.18	68.94	64.39	71.21	64.39	70.45
5	74.80	75.30	73.20	76.50	76.10	72.30
6	83.33	82.96	77.41	80.74	68.15	81.11
7	85.81	83.87	81.94	85.81	84.52	83.23
8	90.11	95.40	92.64	96.09	95.63	96.09
9	80.67	82.24	80.51	82.01	33.31	82.27
10	70.15	98.12	98.33	94.89	97.81	82.57
11	83.19	88.32	86.32	93.45	91.17	92.31
12	68.27	75.48	83.17	80.29	59.62	78.37

Al analizar los resultados obtenidos se determinó que los datos en cada uno de los grupos experimentales no estaban distribuidos normalmente según el Test de Shapiro-Wilk con una significación al 5%. Para determinar la existencia de diferencias significativas entre VD y MI, VM, S, Ba y Bo se utiliza la modificación de Iman-Davenport del Test de Friedmann [20]. El *p.value* obtenido ($<2.2E^{-16}$) muestra que existe al menos un grupo experimental cuyos resultados son diferentes al de los demás. Con el objetivo de establecer entre qué grupo existen las diferencias, se procede a aplicar el test *post-hoc* de Friedman con la corrección de *Bergmann & Hommel's* por sus potencialidades en el análisis de datos [21].

En la Tabla III se puede observar que según los

valores de significación obtenida, existen diferencias significativas entre VD y los restantes grupos experimentales.

Tabla III: Valores de significación corregidos por Bergmann & Hommel's para el Test de Friedman

	MI	VM	S	Ba	Bo
MI	-				
VM	0	-			
S	0	0.179	-		
Ba	0.3049	0	0	-	
Bo	0.0487	0	0	0.179	-
VD	0	0.0005	0.0502	0	6.66E-15

Se omite la matriz triangular superior por ser simétrica a la observada en la tabla. En negritas se señalan los grupos que son diferentes estadísticamente

Aun así, al analizar el promedio de los rangos obtenidos en cada uno de los grupos, se puede observar en la Figura. 3 que el desempeño de VD, S y VM estuvo por debajo de MI, Ba y Bo. En esta figura los nodos conectados no presentan diferencias significativas y el número indica el promedio de los rangos obtenidos para ese grupo [20]. Un valor pequeño indica el mejor desempeño. Como puede observarse, la combinación de clasificadores por voto dinámico tiene un desempeño por detrás del obtenido con mejor clasificador individual, con *Bagging* y con *Boosting*; pero obtiene mejores resultados que sus homólogos *Stacking* y Voto mayoritario, siendo este último el peor de todos.

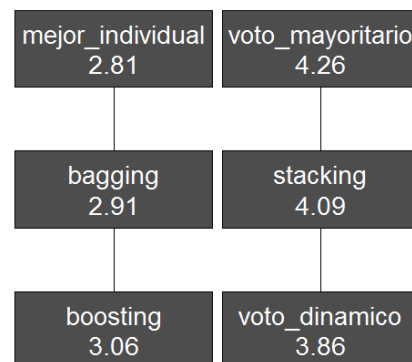


Figura. 3: Grafo de comparación entre los grupos experimentales, según el promedio de los rangos de cada grupo

Finalmente, en la Figura. 4 puede observarse como

la combinación de clasificadores por voto dinámico incrementó la cantidad de combinaciones cuyo desempeño era superior al voto mayoritario. Similares resultados pueden observarse con respecto a *Stacking*, no siendo así para el mejor clasificador individual de la combinación, *Bagging* y *Boosting*.

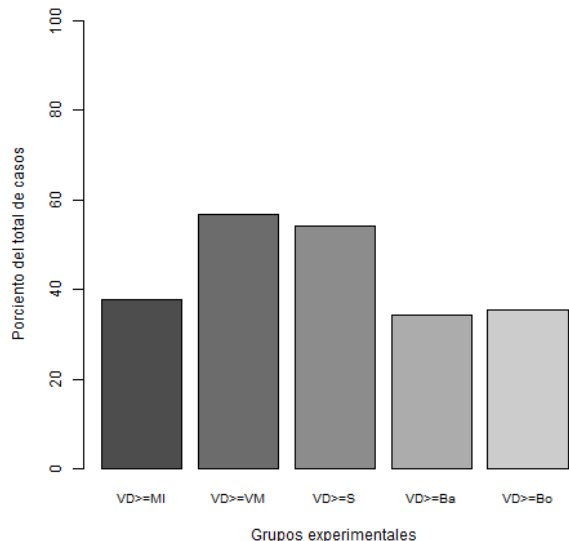


Figura. 4: Comparación entre los resultados obtenidos con VD y MI, VM, Ba y Bo

En un análisis descriptivo de los resultados obtenidos para VM, S y VD se determinó que en un 32% del total casos (222 de 684 casos según la configuración del experimento realizado), la mejor exactitud individual estaba por encima del 90% de casos correctamente clasificados. De este, solamente el 13% de casos se corresponden a combinaciones construidas con voto dinámico que superaban la mejor individual, siendo precedidos por la combinación con voto mayoritario (17%) y por las de *Stacking* (37%), siendo esta opción la mejor de las tres. El resto de los casos se corresponden con los resultados obtenidos en *Bagging* y *Boosting*, no considerados en este análisis. Estos resultados tan bajos para VD están en correspondencia con la penalización propuesta dado que en la combinación de varios clasificadores individuales muy buenos, la penalización calculada a partir de los errores de clasificación no tiene mucha influencia en el resultado final del sistema. Por lo tanto, se recomienda usar la combinación por voto dinámico para solucionar problemas donde el desempeño individual de los clasificadores es muy pobre.

4. CONCLUSIONES

En el presente trabajo se diseña e implementa una modificación al algoritmo de voto mayoritario existente en el WEKA y se aplica una variante del voto pesado descrito por la autora para considerar el

desempeño de los clasificadores bases durante la clasificación.

Se comparan los resultados de VD con respecto al mejor clasificador individual de cada combinación, el voto mayoritario original y otros modelos más populares como *Bagging*, *Boosting* y *Stacking*. El análisis derivó en que, para las bases de casos y los clasificadores individuales utilizados, los resultados de VD superaban significativamente a los obtenidos por voto mayoritario y que muchas de las combinaciones formadas con VD superaban la exactitud obtenida con *Stacking*.

A pesar de estos resultados, se recomienda utilizar la modificación propuesta en la solución de problemas en los que el desempeño de los clasificadores individuales no sea muy bueno para explotar mejor las ventajas de la penalización agregada al esquema de votación.

La modificación propuesta fue implementada en un multiclasificador similar al Vote implementado en el WEKA, e incluido como paquete de esta herramienta para que pueda ser ampliado y utilizado en otros estudios.

5. REFERENCIAS BIBLIOGRÁFICAS

- 1 **Kuncheva, L. I.:** Combining Pattern Classifiers: Methods and Algorithms, 2004.
- 2 **Dietterich, T. G.:** "Ensemble methods in machine learning. Multiple Classifier Systems", 2000.
- 3 **Bonet, I.; A. Rodríguez; M. M. García y R. Grau:** "Combinación de clasificadores para bioinformática", Computación y Sistemas, Vol. 16, No. 2, pp. 191-201, 2012.
- 4 **Witten, I. H.; E. Frank; M. A. Hall y C. J. Pal:** Data Mining: Practical machine learning tools and techniques, 2016.
- 5 **Sigletos, G.; G. Paliouras; C. D. Spyropoulos y M. Hatzopoulos:** "Combining information extraction systems using voting and stacked generalization", Journal of Machine Learning Research, Vol. 6, No. Nov, pp. 1751-1782, 2005.
- 6 **Montero, P. E. F.:** "Modelo de multiclasificador basado en selección de rasgos para predecir la afinidad de antagonistas por receptores de adenosina A2B", Universidad Central "Martha Abreu" de las Villas, Villa Clara, 2011.
- 7 **Asuncion, A. y D. J. Newman:** UCI Machine Learning Repository, [http://www.ics.uci.edu/sim\\$mllearn/MLRepository.html](http://www.ics.uci.edu/sim$mllearn/MLRepository.html)
- 8 **Cunningham, P. y J. Carney:** "Diversity

versus Quality in Classification Ensembles Based on Feature Selection", ECML, pp. 109-116, 2000.

9 **Hernández, L. C.:** "Estudios de medidas de diversidad en sistemas multclasificadores". Maestría en Ciencias de la Computación, Universidad Central "Martha Abreu" de las Villas, Santa Clara, 2013.

10 **Kuncheva, L. I. y C. J. Whitaker:** "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy", Machine Learning, Vol. 51, pp. 181-207, 2003.

11 **Kuncheva, L. I. y L. C. Jain:** "Designing classifier fusion systems by genetic algorithms", IEEE Transactions on Evolutionary Computation, Vol. 4, No. 4, pp. 327-336, 2000.

12 **Martínez, L. R. S.:** "Construcción de sistemas multclasificadores usando Optimización de Colonias de Hormigas". Licenciatura en Ciencias de la Computación, Universidad Central "Marta Abreu" de Las Villas, 2015.

13 **Morales, A.:** "Construcción de sistemas multclasificadores usando Algoritmos Genéticos y medidas de diversidad". Licenciatura en Ciencias de la Computación, Universidad Central "Marta Abreu" de Las Villas, 2014.

14 **Palanisamy, S. y S. Kanmani:** "Classifier Ensemble Design using Artificial Bee Colony based Feature Selection", International Journal of Computer Science Issues (IJCSI), Vol. 9, No. 3, 2012.

15 **Breiman, L.:** "Bagging predictors", Machine Learning, pp. 123-140, 1996.

16 **Breiman, L.:** "Random forests", Machine Learning, Vol. 45, No. 1, pp. 5-32, 2001.

17 **Woźniak, M.; M. Graña y E. Corchado:** "A survey of multiple classifier systems as hybrid systems", Information Fusion, Vol. 16, pp. 3-17, 2014.

18 **Arlot, S. y A. Celisse:** "A survey of cross-validation procedures for model selection", Statistics surveys, Vol. 4, pp. 40-79, 2010.

19 **Lavesson, N.:** "Evaluation and analysis of supervised learning algorithms and classifiers",

Blekinge Institute of Technology, 2006.

20 **Demsar, J.:** "Statistical Comparisons of Classifiers over Multiple Data Sets", Journal of Machine Learning Research, Vol. 7, pp. 1-30, 2006.

21 **García, S. y F. Herrera:** "An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for All Pairwise Comparisons", Journal of Machine Learning Research, Vol. 9, pp. 2677-2594, 2008.

6. SÍNTESIS CURRICULARES DE LOS AUTORES



Alejandro Morales Hernández Lic. en Ciencias de la Computación en la Universidad Central "Marta Abreu" de Las Villas (UCLV), Cuba (2014). Profesor Instructor del Departamento de Inteligencia Artificial del Centro de Investigaciones Informáticas de la UCLV. Experiencia profesional en Inteligencia Artificial, Medidas de Diversidad en sistemas multclasificadores, Reconocimiento de patrones y Estadística. Áreas de interés: Inteligencia Artificial y Bioinformática. Correo electrónico personal: alejandro.morales2014@yandex.com. Dirección postal: Calle H #18 E/ 3ra y 9na Reparto Guamajal, Santa Clara, Villa Clara



Claudia Fernández Hernández Lic. en Ciencias de la Computación en la Universidad Central "Marta Abreu" de Las Villas (UCLV), Cuba (2014). Especialista superior en TICs en Empresa de Tecnologías de la Información para la Defensa, XETID. Experiencia profesional en Inteligencia Artificial y desarrollo para sistemas GNU Linux. Áreas de interés: Inteligencia Artificial y Optimización



Leidys Cabrera Hernández Lic. en Ciencias de la Computación en la Universidad Central "Marta Abreu" de Las Villas (UCLV), Cuba (2009). Máster en Bioinformática y Biología Computacional, UCLV (2013). Profesora Asistente del Departamento de Inteligencia Artificial del Centro de Investigaciones Informáticas de la UCLV. Experiencia profesional en Inteligencia Artificial y Bioinformática. Áreas de interés: Inteligencia Artificial