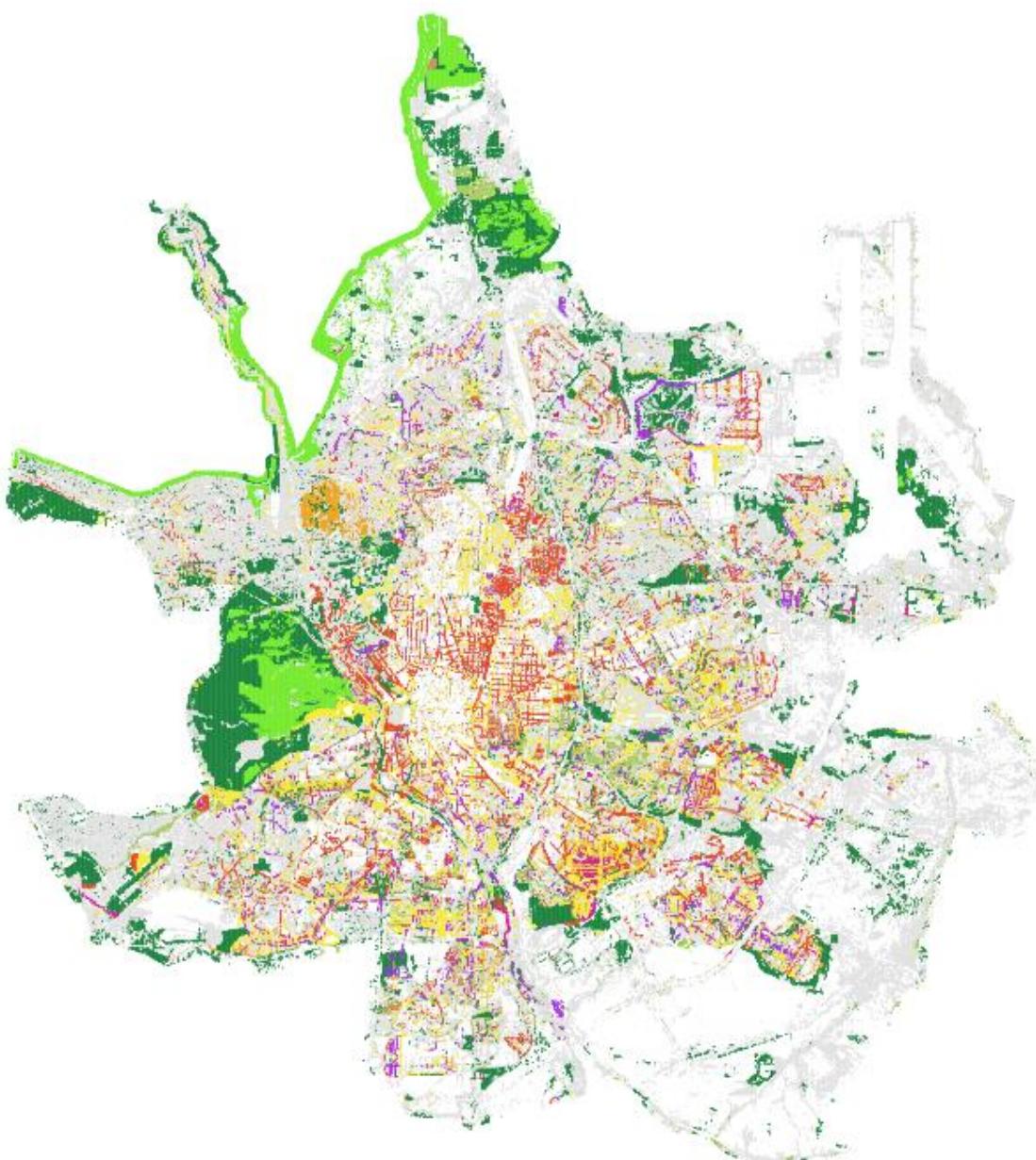


Centro de Observación y
Teledetección Espacial,
S.A.U.

(COTESA – A47461066)



SERVICIO DE CARTOGRAFÍA DE ESPECIES

[Metodología para la detección de especies en el segundo semestre de 2024 sobre el término municipal de Madrid más las pistas del aeropuerto Adolfo Suárez – Madrid Barajas sin el Monte del Pardo.]

Contenido

1.	Introducción	2
2.	Objetivo.....	2
3.	Metodología	2
3.1.	Cálculo de índices.....	2
3.2.	Selección de áreas de entrenamiento.....	4
3.3.	Clasificación por píxel	5
3.4.	Resultados.....	6
3.5.	Entregables.....	8

1. Introducción.

En este informe se muestra la metodología utilizada para la obtención de especies en Madrid para el segundo semestre del año 2024. Para ello se ha utilizado un algoritmo propio de COTESA que permite la automatización de cartografía de coberturas de suelo de alta resolución. Es una herramienta diseñada en software libre, que permite la semi- automatización de la cartografía de ocupación del suelo y con mayor precisión y recurrencia que la cartografía existente actual, estos algoritmos han sido modificados para así poder determinar qué especies pueden ser detectadas.

Mediante la clasificación supervisada basada en píxeles del mosaico obtenido anteriormente, se logra una cartografía de coberturas basadas en la detección de especies, que son 10 clases genéricas y una que engloba a otras especies, teniendo así: Pino, Falso Platanero, Arce, Olmo, Acacia, Falsa Acacia, Almez, Aligustre, Castaño, Encina y otros (resto de especies arbóreas). Se automatiza así el cartografiado de especies arbóreas que serán englobadas en teselas de 20x20 metros donde prevalecerá la especie con mayor superficie.

2. Objetivo.

Obtención de una capa shapefile y raster donde queden reflejadas las especies arbóreas para el segundo semestre de 2024, para el término municipal de Madrid (más pistas del Aeropuerto Adolfo Suárez - Madrid Barajas).

Aunque se utilicen índices espectrales de verano se consideran del segundo semestre ya que las bandas espectrales son las de la imagen de otoño, es decir las bandas de otoño más los índices de la misma fecha a la que se añaden los índices de verano para obtener más información.

3. Metodología.

A continuación, se describe cada una de las fases llevadas a cabo para la obtención de las especies, que radica en varios procesos.

3.1. Cálculo de índices.

El cálculo de índices corresponde a la preparación de las imágenes para poder realizar operaciones de segmentación y clasificación.

El uso de índices espectrales para facilitar la clasificación de la imagen procede de dos índices diferentes, NDVI y RVI. Con estos índices adquirimos información explícita sobre vegetación y agua que favorecerá la clasificación de estas zonas.

- NDVI (Índice de Vegetación de Diferencia Normalizada): es un índice usado para estimar la cantidad, calidad y desarrollo de la vegetación con base a la medición (por medio de sensores remotos instalados comúnmente en una plataforma espacial) de la intensidad de la radiación de ciertas bandas del espectro electromagnético que la vegetación emite o refleja.

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

- RVI (Índice Relativo de Vigor): es una relación entre la reflectancia registrada en las bandas de infrarrojo cercano (NIR) y rojo. Esta es una forma rápida de distinguir las hojas verdes de otros objetos en la escena y estimar la biomasa relativa presente en la imagen. Además, este valor puede ser muy útil para distinguir la vegetación estresada de las áreas no estresadas.

$$RVI = \frac{NIR}{RED}$$



Imagen 1. NDVI a la izquierda, RVI a la derecha.

La combinación de los resultados de ambos índices se une, como si de dos bandas más se tratase al mosaico original y aportando así una información extra al mismo.

Esta operación se repetirá con la imagen de verano de 2024, con ello conseguiremos tener dos imágenes diferentes con 2 índices cada una, es decir tendremos una sola imagen con 12 bandas (RGB-NIR-NDVI-RVI del primer semestre de 2024 y RGB-NIR-NDVI-RVI de segundo semestre de 2024), esta variabilidad temporal permitirá tener un mayor volumen de datos por especie, favoreciendo así la clasificación de las mismas.

3.2. Selección de áreas de entrenamiento.

Las áreas de entrenamiento se obtienen mediante una capa de especies arbóreas entregada por cliente. Esta capa, de puntos, es la utilizada como entrada para el algoritmo, como se puede observar en la Imagen 2.

Es en este paso en el que, en función del número de muestras, así como de la representatividad de las mismas se decide qué especies pueden ser las más idóneas para que sean detectadas, teniendo en cuenta varios aspectos, como la estacionalidad, la respuesta espectral, etc. Y de las 769.293 muestras, se observa que las que mayor probabilidad tienen de ser detectadas por sus características son las 10 clases previamente dichas.

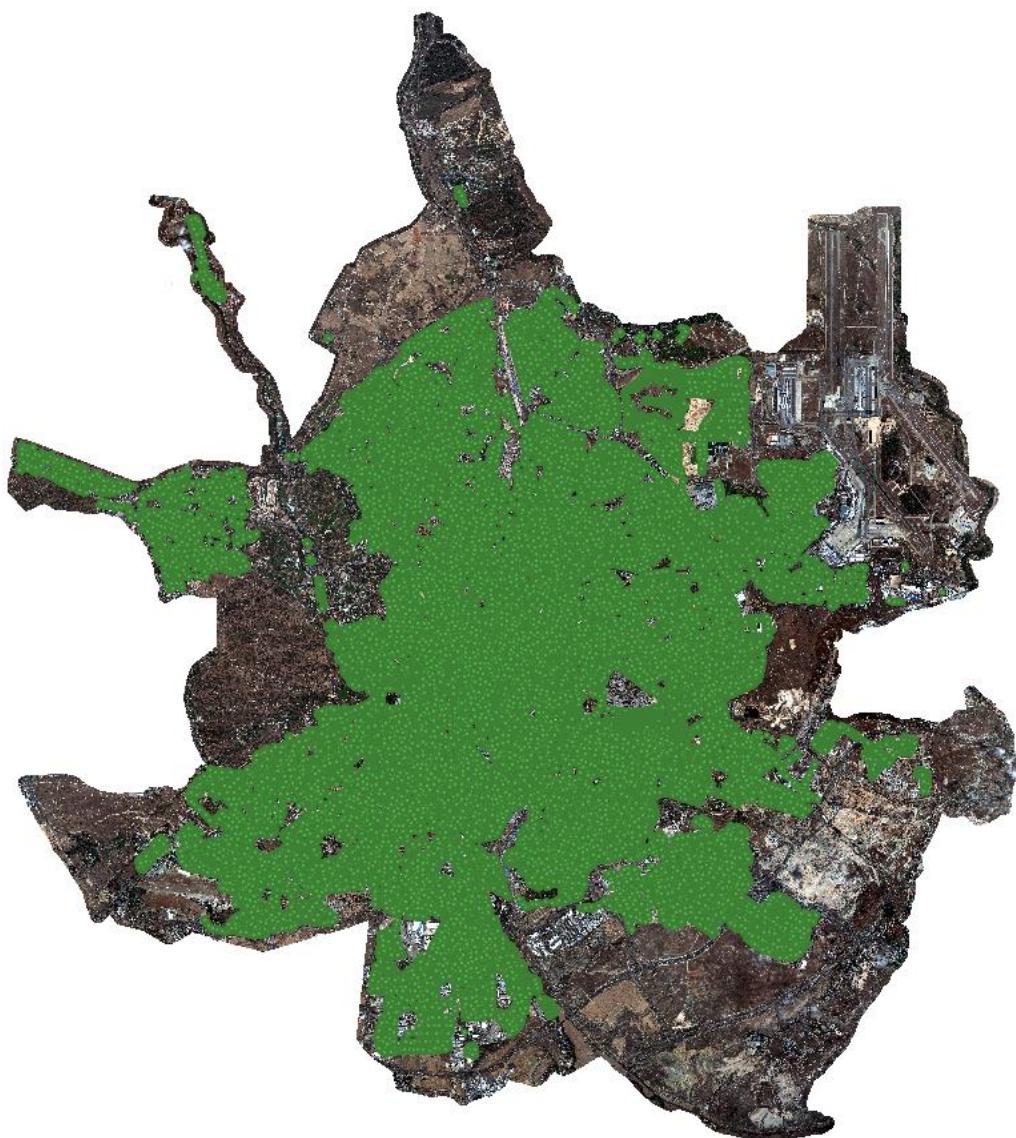


Imagen 2. Puntos utilizados como áreas de entrenamiento para la detección de especies.

3.3. Clasificación por píxel.

La clasificación se realizará con los puntos obtenidos de la capa que fue provisionada, y a diferencia que en la detección de zonas verdes (donde las áreas de entrenamiento eran polígonos en los que se creaban puntos aleatorios y posteriormente se filtraban) no se filtrarán los puntos más que por las especies a detectar, salvo la clase otros donde se engloban todas aquellas especies que no van a poder ser detectadas individualmente. De estos puntos serán extraídas las estadísticas de las diferentes bandas del mosaico (recordad que cuenta con 12 bandas diferentes), siendo estos datos con los que el algoritmo entrenará obteniendo un modelo de clasificación.

Tras esto se utilizarán las siguientes herramientas:

- *ComputeImageStatistics*.
- *TrainVectorClassifier* donde se elegirá el método de entrenamiento de clasificación que es el *Random Forest*.
- *ImageClassifier* que utilizando las estadísticas previas clasificará la imagen.

Tras ello se realizan una serie de procesos para obtener los segmentos, siendo éstos un filtrado de los pixels para así poder depurar la imagen omitiendo el ruido que pueda haber en la clasificación mediante la herramienta de *Filtrado*, seguido de la utilización de la herramienta de ráster a polígono (*poligonizar*), y finalmente la herramienta *dissolve* de las clases obtenidas, como se puede observar en la imagen 3.

original	0	1	2	3	4	5	6	7	8	9
7480	675,0000000000...	868,0000000000...	1371,0000000000...	1773,0000000000...	1893,0000000000...	2336,0000000000...	2484,0000000000...	2577,0000000000...	0,136995673179...	-0,29655628180...
7483	728,0000000000...	853,0000000000...	1490,0000000000...	1921,0000000000...	2081,0000000000...	2356,0000000000...	2366,0000000000...	2275,0000000000...	0,061623669474...	-0,23597467334...
7484	531,0000000000...	598,0000000000...	1189,0000000000...	1578,0000000000...	1751,0000000000...	1893,0000000000...	1914,0000000000...	1812,0000000000...	0,044474761833...	-0,23364485800...
7488	477,0000000000...	599,0000000000...	975,0000000000...	1267,0000000000...	1407,0000000000...	1621,0000000000...	1767,0000000000...	1837,0000000000...	0,113421551883...	-0,28884026400...
7514	342,0000000000...	555,0000000000...	976,0000000000...	1870,0000000000...	2115,0000000000...	2988,0000000000...	3148,0000000000...	3484,0000000000...	0,22570168801...	-0,54856616238...
7520	81,0000000000...	188,0000000000...	284,0000000000...	322,0000000000...	368,0000000000...	803,0000000000...	1284,0000000000...	1315,0000000000...	0,5486079634723...	-0,633074893825...
7523	548,0000000000...	731,0000000000...	1128,0000000000...	1471,0000000000...	1598,0000000000...	1928,0000000000...	2141,0000000000...	2197,0000000000...	0,140838901400...	-0,30988070360...
7524	477,0000000000...	895,0000000000...	990,0000000000...	1391,0000000000...	1501,0000000000...	1918,0000000000...	2216,0000000000...	2411,0000000000...	0,192359432578...	-0,79546599884...
7527	181,0000000000...	331,0000000000...	558,0000000000...	698,0000000000...	823,0000000000...	1595,0000000000...	2374,0000000000...	2481,0000000000...	0,495142326394...	-0,61937247746...
7529	294,0000000000...	485,0000000000...	687,0000000000...	1031,0000000000...	1174,0000000000...	1543,0000000000...	1779,0000000000...	1863,0000000000...	0,204876595879...	-0,43899514805...
7530	353,0000000000...	644,0000000000...	1011,0000000000...	1663,0000000000...	1951,0000000000...	2468,0000000000...	2775,0000000000...	2924,0000000000...	0,174059967369...	-0,46649077334...
7531	351,0000000000...	511,0000000000...	790,0000000000...	1091,0000000000...	1177,0000000000...	1799,0000000000...	2286,0000000000...	2356,0000000000...	0,320242553849...	-0,48634591898...
7532	182,0000000000...	633,0000000000...	938,0000000000...	1251,0000000000...	1386,0000000000...	1975,0000000000...	2419,0000000000...	2590,0000000000...	0,271484881639...	-0,4116771221...
7533	247,0000000000...	442,0000000000...	781,0000000000...	1038,0000000000...	1203,0000000000...	2258,0000000000...	3281,0000000000...	3398,0000000000...	0,46342517882...	-0,62345373630...
7534	215,0000000000...	410,0000000000...	748,0000000000...	1163,0000000000...	1351,0000000000...	2426,0000000000...	3257,0000000000...	3565,0000000000...	0,413628458876...	-0,62546669291...
7537	386,0000000000...	602,0000000000...	1031,0000000000...	2056,0000000000...	2965,0000000000...	3197,0000000000...	3487,0000000000...	3477,0000000000...	0,15234627930...	-0,5418036865...
7539	375,0000000000...	621,0000000000...	1077,0000000000...	2198,0000000000...	2795,0000000000...	3438,0000000000...	3642,0000000000...	3761,0000000000...	0,142229884862...	-0,5435473128...
7542	164,0000000000...	256,0000000000...	449,0000000000...	846,0000000000...	977,0000000000...	1285,0000000000...	1447,0000000000...	1387,0000000000...	0,193894386291...	-0,52817130623...
7546	201,0000000000...	319,0000000000...	664,0000000000...	1081,0000000000...	1664,0000000000...	2164,0000000000...	2196,0000000000...	2362,0000000000...	0,150995986139...	-0,55135126842...
7588	408,0000000000...	879,0000000000...	1121,0000000000...	1951,0000000000...	2324,0000000000...	3272,0000000000...	3881,0000000000...	3942,0000000000...	0,250925673412...	-0,55177162721...
7589	180,0000000000...	281,0000000000...	487,0000000000...	852,0000000000...	1143,0000000000...	1477,0000000000...	1543,0000000000...	1552,0000000000...	0,140865477809...	-0,553523381842...
7614	1157,0000000000...	1438,0000000000...	1307,0000000000...	871,0000000000...	764,0000000000...	1147,0000000000...	1685,0000000000...	1848,0000000000...	0,378809180450...	-0,12924717386...
7615	1496,0000000000...	1776,0000000000...	1402,0000000000...	862,0000000000...	861,0000000000...	765,0000000000...	1114,0000000000...	1185,0000000000...	0,241225670644...	0,15091167424...
7661	588,0000000000...	864,0000000000...	828,0000000000...	903,0000000000...	966,0000000000...	1054,0000000000...	1078,0000000000...	1124,0000000000...	0,03479420139...	-0,13116474448...
5967	45,0000000000...	51,0000000000...	150,0000000000...	169,0000000000...	177,0000000000...	352,0000000000...	568,0000000000...	681,0000000000...	0,524832189883...	-0,38217265192...
5068	13,0000000000...	49,0000000000...	116,0000000000...	115,0000000000...	107,0000000000...	461,0000000000...	964,0000000000...	1091,0000000000...	0,803649597113...	-0,78809093141...
5069	14,0000000000...	72,0000000000...	145,0000000000...	138,0000000000...	136,0000000000...	704,0000000000...	1476,0000000000...	1564,0000000000...	0,3239946300497...	-0,61997531852...
5070	12,0000000000...	59,0000000000...	123,0000000000...	119,0000000000...	545,0000000000...	1161,0000000000...	1233,0000000000...	1214,0000000000...	0,314062476158...	-0,60139642953...

Imagen 3. Datos obtenidos de los diferentes puntos sobre 10 de las 12 bandas de la imagen.

3.4. Resultados.

En cuanto a los resultados obtenidos, una vez realizada la clasificación por pixel, se vio que, en determinadas especies debido sobre todo a posicionamientos de sombra sobre las copas de los árboles, podían ser clasificados de diferentes especies un mismo árbol, es por ello que para solventarlo, se establecen teselas de 20x20 metros donde la especie que más superficie ocupe será la predominante.

Para la realización del QC se utilizará la matriz de confusión y el índice Kappa.

La matriz de confusión nos permite comparar el resultado de la clasificación con el conjunto de datos de referencia. Las muestras de referencia/validación están en filas mientras que los datos de clasificación están en columnas. Es equivalente a una matriz de contingencia. Los elementos diagonales corresponden a píxeles (o áreas) correctamente clasificados, mientras que todos los otros valores son clasificaciones erróneas, llamados confusiones.

	1	2	3	4	5	6	7	8	9	10	11
1	12234	187	93	48	234	59	128	85	33	634	145
2	334	34233	198	92	287	173	108	29	127	52	136
3	45	88	10767	25	152	26	18	61	52	49	368
4	38	182	115	18868	129	75	62	26	45	43	35
5	16	24	42	56	19497	188	42	119	38	17	54
6	25	15	38	19	227	9531	39	18	12	16	41
7	289	254	124	157	148	107	10911	41	32	62	79
8	33	59	78	31	218	35	38	8525	27	23	53
9	24	116	58	21	38	39	11	13	5662	10	42
10	20	11	9	14	10	14	11	11	22	739	15
11	7483	11927	35145	315	924	325	756	1278	451	1843	469503

Tabla 1. Matriz de confusión donde 1 representa la especie de pinos, 2 falso platanero, 3 arces, 4 olmos, 5 acacias, 6 falsas acacias, 7 almeces, 8 aligustres, 9 castaños, 10 encinas, 11 otros.

Con la realización de la matriz de confusión, se crean a su vez una serie de índices muy orientativos respecto a la clasificación. Como se puede observar en la tabla 1, aparecen representados todos los puntos extraídos con la herramienta *sample extraction*, y son estos los utilizados para realizar la clasificación, teniendo en las columnas y las filas las diferentes clases y cuánto ha variado la clasificación entre ellas, observando una amplia mayoría de puntos bien clasificados en función a la clasificación previa dada como área de entrenamiento.

Este primer índice ofrece una visión general de la precisión global de la clasificación. No refleja la diferencia del número de muestras entre clases. El índice Kappa se extraerá en función a las áreas de entrenamiento hechas previamente, ya que son de una alta fiabilidad porque fueron hechas a mano por un fotointérprete.

El índice de kappa responde más eficazmente a este desequilibrio frecuente, utilizando un ajuste estadístico. En ambos casos el resultado es un valor comprendido entre 0 y 1, cuanto más se acerca a 1 más eficiente es la clasificación.

El índice Kappa:

$$\kappa = \frac{N \sum_{i=1}^m x_{ii} - \sum_{i=1}^m x_{i\sum} x_{\sum i}}{N^2 - \sum_{i=1}^m x_{i\sum} x_{\sum i}}$$

m = número total de clases.

N = número total de píxeles en las m clases de referencia.

x_{ii} = elementos de la diagonal de la matriz de confusión.

$x_{i\sum}$ = suma de los píxeles de la clase i de referencia.

$x_{\sum i}$ = suma de los píxeles clasificados como la clase i .

Para ello, se utilizarán como referencia los puntos que se obtuvieron en el proceso de clasificación por píxel, en concreto tras el paso de la utilización de la herramienta de *sample extraction* con la que se cribaron los puntos obtenidos aleatoriamente dentro de las propias áreas de entrenamiento. Con ello veremos la bondad del algoritmo en esta imagen comparando las áreas de entrenamiento con los resultados obtenidos.

<i>Especie detectada</i>	<i>Detección Total</i>	<i>Detectado</i>	<i>Porcentaje de acierto</i>
Pino (Pinus pinea)	14122	12234	86.63%
Plataneros (Platanus hispanica)	35712	34233	95.85%
Arces (Acer monspessulanum)	11495	10767	93.66%
Olmos (Ulmus minor)	19697	18868	95.79%
Acacias (Albizia julibrissin)	20109	19497	96.96%
Falsas acacias (Robinia pseudoacacia)	9965	9531	95.65%
Almeces (Celtis australis)	12243	10911	89.12%
Aligustres (Ligustrum lucidum)	9089	8525	93.80%
Castaños (Castanea sativa)	5844	5662	96.89%
Encinas (Quercus ilex)	846	739	87.35%
Resto de especies	528430	469503	88.85%

Tabla 2. Especies detectadas, tras el QC y detectadas automáticamente, con el porcentaje de acierto.

Pueden parecer valores muy elevados, pero siempre hay que tener en cuenta, que este tipo de evaluación se hace siempre sobre los propios datos de entrada, es decir, de las áreas de entrenamiento, áreas que han sido escogidas y delimitadas de manera muy cautelosa.

3.5. Entregables.

A continuación, se muestran los entregables con una breve explicación de los mismos.

- *EsAytMadridOt2024VegTipEsp.tif*; detección de especies con el algoritmo propio de COTESA en formato raster.
- *EsAytMadridOt2024VegTipEsp_tif.xml*; metadatos del archivo raster.
- *EsAytMadridVe2024VegTipEsp.shp*; detección de especies con el algoritmo propio de COTESA en formato shapefile.
- *EsAytMadridOt2024VegTipEsp_shp.xml*; metadatos del archivo shapefile.
- *EsAytMadridOt2024VegInf.pdf*; documentación con los procedimientos seguidos en la creación de la detección de especies.