

CAPÍTULO 6.

RESULTADOS Y CONCLUSIONES.

6.1 PROCEDIMIENTO. PARÁMETROS DE CALIDAD.

6.1.1 Procedimiento.

Aunque a lo largo de la memoria de este proyecto se ha ido describiendo con precisión cuáles han sido los pasos seguidos, exponemos, ahora que se ha adquirido una visión de conjunto, un esquema de nuestra manera de proceder.

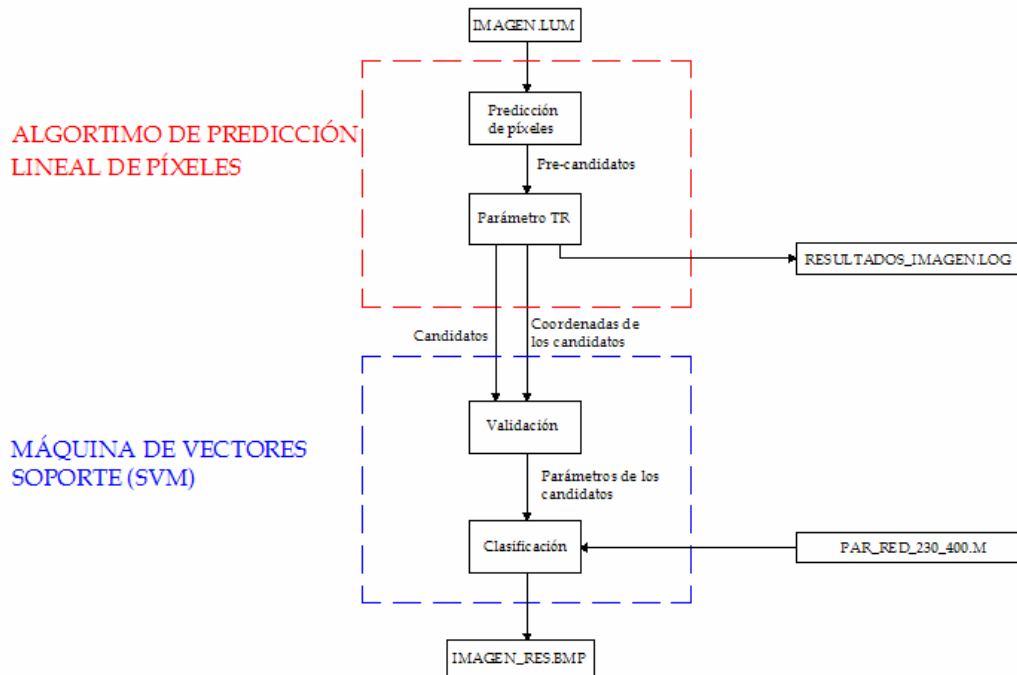


Figura 6.1. Esquema general de funcionamiento.

En efecto, partimos de una imagen en formato *LUM* y obtenemos otra en formato *BMP* con las microcalcificaciones señaladas. Generamos también un fichero de salida en formato *LOG* que almacena las posibles incidencias del proceso, así como los datos estadísticos del mismo.

Para la aplicación correspondiente a la máquina de vectores soporte, resulta imprescindible tener en el mismo directorio el fichero que utilizamos para entrenar a la red, *par_red_230_400.m*.

En este apartado vamos a definir ciertos parámetros que usaremos como indicadores de la validez de nuestro trabajo. Asimismo, compararemos dichos resultados con los que teníamos en la versión anterior, veremos cuáles han sido las mejores introducidas y discutiremos sobre las limitaciones que posee.

6.1.2 Parámetros de calidad.

No todas las microcalcificaciones existentes son detectadas, ni todas las detectadas lo son verdaderamente. Tenemos, pues, cuatro *posibles casos* de microcalcificaciones:

1. **Verdaderos positivos (VP):** son el número de microcalcificaciones que han sido detectadas correctamente.
2. **Falsos positivos (FP):** son aquellos píxeles que han sido considerados como centro de una microcalcificación pero que, en realidad, no lo son.
3. **Falsos negativos (FN):** son aquellas microcalcificaciones reales que no han sido detectadas.
4. **Verdaderos negativos (VN):** son aquellos candidatos que han sido bien descartados por la SVM.

La medicina es una ciencia de probabilidades y un arte de manejar la incertidumbre. Dicha incertidumbre se extiende no sólo a las actividades preventivas y terapéuticas sino también a las diagnósticas. Es evidente que una buena prueba diagnóstica es la que ofrece resultados positivos en enfermos y negativos en sanos. Por lo tanto, las *condiciones* que deben ser exigidas a una prueba son:

1. **Validez:** Es el grado en que una prueba mide lo que se supone que debe medir. La *sensibilidad* y la *especificidad* de un análisis son medidas de su validez. Definémoslas:
 - a. **Sensibilidad.** Es la probabilidad de clasificar correctamente a una microcalcificación, es decir, la probabilidad de que para una microcalcificación se obtenga en la prueba un resultado positivo.

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

- b. **Especificidad.** Es la proporción de candidatos que han sido bien descartados por el programa. Su fórmula es:

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

2. **Reproductividad:** es la capacidad de la prueba para ofrecer los mismos resultados cuando se repite su aplicación en circunstancias similares. La variabilidad biológica del hecho observado y la introducida por el propio observador, determinan su reproductividad.
3. **Seguridad:** La seguridad viene determinada por el valor predictivo de un resultado positivo o negativo. ¿Con qué seguridad una prueba predecirá la presencia o ausencia de microcalcificación? Y ante un resultado positivo, ¿qué probabilidad existe de que este resultado indique presencia de la microcalcificación? Tanto la sensibilidad como la especificidad proporcionan información acerca de la probabilidad de obtener un resultado concreto (positivo o negativo) en función de la verdadera condición del enfermo con respecto a la enfermedad. Sin embargo, cuando a un paciente se le realiza alguna prueba, el médico carece de información *a priori* acerca de su verdadero diagnóstico, y más bien la pregunta se plantea en sentido contrario: ante un resultado positivo (o negativo) en la prueba, ¿cuál es la probabilidad de que el paciente esté realmente enfermo (o sano)?. Así pues, resulta obvio que hasta el momento sólo hemos abordado el problema en una dirección. Por medio de los valores predictivos completaremos esta información:

- a. **Valor predictivo positivo (PPV).** Es la probabilidad de que exista una microcalcificación habiéndose obtenido un resultado positivo. El valor predictivo positivo puede estimarse, por tanto:

$$PPV = \frac{VP}{VP + FP}$$

- b. **Valor predictivo negativo (NPV).** Es la probabilidad de que una microcalcificación descartada en la prueba no esté realmente presente. Se estima:

$$NPV = \frac{VN}{VN + FN}$$

Nosotros usaremos estos cuatro parámetros (sensibilidad, especificidad, PPV y NPV) para determinar la calidad de nuestras detecciones.

6.2 RESULTADOS.

6.2.1 Tabla de resultados.

Mostramos, a continuación, la relación completa de las mamografías objetos de estudio, así como los resultados obtenidos.

Identificador	Pre-candidatos	Candidatos	Micros	VP	FP	FN	VN	Sens.	PPV	Espec.	NPV
c5_96rc	8171	1211	252	252	0	30	929	89,36%	100,00%	100,00%	96,87%
c5_96ro	8823	1514	366	365	1	23	1125	94,07%	99,73%	99,91%	98,00%
c7_91lo	11376	1052	44	42	2	2	1006	95,45%	95,45%	99,80%	99,80%
c16_93lc	4786	476	72	70	2	1	403	98,59%	97,22%	99,51%	99,75%
c16_93rc	5709	444	52	52	0	0	392	100,00%	100,00%	100,00%	100,00%
c16_93ro	8447	818	88	81	7	4	726	95,29%	92,05%	99,05%	99,45%
c16_95lc	5345	888	67	67	0	2	819	97,10%	100,00%	100,00%	99,76%
c16_99rc	6632	633	89	66	23	0	544	100,00%	74,16%	95,94%	100,00%
c52_97rc	6309	856	48	48	0	4	804	92,31%	100,00%	100,00%	99,50%
c52_97ro	8361	1067	48	40	8	4	1015	90,91%	83,33%	99,22%	99,61%
c61_97lo	5034	499	51	50	1	5	443	90,91%	98,04%	99,77%	98,88%
c70_91rc	6032	635	52	47	5	3	580	94,00%	90,38%	99,15%	99,49%
c91_92rc	8199	832	86	86	0	1	745	98,85%	100,00%	100,00%	99,87%
c98_95lo	8513	743	120	116	4	4	619	96,67%	96,67%	99,36%	99,36%
c99_93ro	14481	1374	118	115	3	3	1253	97,46%	97,46%	99,76%	99,76%
c99_95lc	8890	1422	70	68	2	3	1349	95,77%	97,14%	99,85%	99,78%
c99_95lo	12719	1694	132	132	0	7	1555	94,96%	100,00%	100,00%	99,55%
c99_95ro	15048	1373	62	62	0	4	1307	93,94%	100,00%	100,00%	99,69%
c100_92lc	6199	411	86	81	5	4	321	95,29%	94,19%	98,47%	98,77%
c100_92lo	11420	1020	74	68	6	3	943	95,77%	91,89%	99,37%	99,68%
c100_95lo	8755	1031	63	62	1	2	966	96,88%	98,41%	99,90%	99,79%
c100_97lc	8816	710	139	137	2	4	567	97,16%	98,56%	99,65%	99,30%
c102_93lo	11995	1375	78	78	0	1	1296	98,73%	100,00%	100,00%	99,92%
c103_93lc	6010	547	83	82	1	3	461	96,47%	98,80%	99,78%	99,35%
c103_93rc	7392	599	89	87	2	2	508	97,75%	97,75%	99,61%	99,61%
c103_93ro	9412	730	78	76	2	1	651	98,70%	97,44%	99,69%	99,85%
c104_93lc	5140	541	155	155	0	8	378	95,09%	100,00%	100,00%	97,93%
c104_93lo	4551	681	120	120	0	19	542	86,33%	100,00%	100,00%	96,61%
c104_93rc	4484	749	180	179	1	5	564	97,28%	99,44%	99,82%	99,12%
c104_93ro	5192	502	105	104	1	6	391	94,55%	99,05%	99,74%	98,49%
c113_95rc	6371	850	111	111	0	3	736	97,37%	100,00%	100,00%	99,59%
c113_95ro	10101	853	121	119	2	4	728	96,75%	98,35%	99,73%	99,45%
c157_92lc	6282	756	141	141	0	1	614	99,30%	100,00%	100,00%	99,84%
c157_92lex	6391	691	101	101	0	3	587	97,12%	100,00%	100,00%	99,49%
c157_92lo	5564	831	265	263	2	4	562	98,50%	99,25%	99,65%	99,29%
c157_97lc	8229	3216	83	83	0	3	3130	96,51%	100,00%	100,00%	99,90%
TOTALES	285179	33624	3889	3806	83	176	29559				
PROMEDIOS	7921,6389	934	108,03	105,7	2,31	4,89	821,1	95,87%	97,08%	99,63%	99,31%

6.2.2 Comparación de resultados.

Mostramos una tabla resumen, que compara los resultados obtenidos con la versión actual con la versión inicial que disponíamos.

Tabla 6.1. Comparación de resultados entre la versión anterior y la actual.

VERSIÓN	SENSIBILIDAD	ESPECIFICIDAD	PPV	NPV	CANDIDATOS	TIEMPO ¹
ANTERIOR	93.31 %	98.67 %	89.15 %	99.31 %	1094	1'23"
NUEVA	95.87 %	99.63 %	97.08 %	99.31 %	934	28'' ²

6.2.2.1 Sensibilidad.

Se ha conseguido mejorar la sensibilidad en un 2.56%. En principio, este valor podría parecer no demasiado elevado, pero debemos señalar que la aplicación de la que partíamos ya poseía una alta sensibilidad. No obstante, si miramos el resto de porcentajes de los parámetros, vemos que éste es, quizás el que más haya que mejorar. Con todo, tener casi un 96% de sensibilidad parece un resultado más que aceptable.

6.2.2.2 Especificidad.

Tener un porcentaje de un 99.63% en el apartado referente a la especificidad implica que el criterio de cribado de nuestra aplicación, en concreto la parte correspondiente al SVM, funciona casi a la perfección. En futuras investigaciones, este parámetro no deberá ser modificado.

6.2.2.3 Valor predictivo positivo (PPV).

Sin duda alguna, este es el parámetro que más se ha visto beneficiado por nuestra intervención. Anteriormente se tenía un resultado inaceptable, puesto que ni siquiera llegaba a un 90%. Aumentar en casi 10 puntos porcentuales este parámetro, hace que hayan desaparecido prácticamente las microcalcificaciones detectadas erróneamente por nuestro programa.

¹ Las pruebas han sido realizadas en un ordenador con procesador AMD 3.2 GHz y 1 GB de memoria RAM. Estos valores variarán mucho en función del equipo que se disponga.

² Nótese que este valor se ha calculado teniendo en cuenta el tiempo que se tarda en crear el fichero de resultados estadísticos que, en realidad, puede suprimirse si se quiere una mayor velocidad de ejecución. En este caso, el tiempo medio se reduce a 23".

6.2.2.4 Valor predictivo negativo (NPV).

El valor de este parámetro ha permanecido, casualmente, al mismo valor. Consideramos que obtener un 99.31% es un excelente resultado.

6.2.2.5 Candidatos.

El valor al que hacemos referencia en la tabla, es el valor medio de candidatos por mamografía. Hemos disminuido en 160 dicho número, lo que supone una reducción del 14.63%. En definitiva, hemos conseguido eliminar candidatos y conseguir mejores resultados, lo que supone que hemos seleccionado candidatos de *más calidad*.

6.2.2.6 Tiempo medio de ejecución.

Sin duda, la mejora más significativa de nuestro programa estriba en lo reducido de su tiempo de ejecución. Hemos pasado de 83" a 23", esto es, hemos reducido 3.61 veces el tiempo de ejecución. Las razones que han dado lugar a este fenómeno pueden resumirse en:

1. Se han eliminado numerosos *bucles for* redundantes, provocando de este modo una optimización de recursos.
2. Eliminamos el análisis del fondo de la imagen. Esto lo hemos conseguido imponiendo una condición: si el valor medio de gris del bloque que estamos analizando es inferior a 50 en formato BMP (lo que equivale a ser mayor de 3600 en formato LUM) consideramos que estamos en una zona negra. Hemos dejado un margen aceptable, ya que hasta 65 podría considerarse como fondo.
3. En la versión anterior se escribían los valores de los vectores de error en un fichero de salida. Este hecho ralentizaba mucho la ejecución. En contrapartida, hemos creado un fichero que aumenta en 5" el tiempo medio de procesado pero que consideramos de gran utilidad ya que almacena valores fundamentales a la hora de analizar los resultados obtenidos.
4. Del mismo modo, como ya se ha comentado en capítulos anteriores, la creación del parámetro *UMBRALMAMA*, utilizado para calcular la

media de gris global de la imagen, también ayuda a bajar el tiempo medio. Tengamos en cuenta que estamos trabajando de imágenes de alrededor de 55 MB de tamaño y con este umbral sólo recorreremos la imagen al leerla en un 21% (en promedio), es decir, nos ahorramos gestionar 44 MB durante su procesado.

6.2.3 Consideraciones generales.

Haciendo un análisis global de los resultados obtenidos, se observa que sólo en dos mamografías (*c5_96rc* y *c5_96ro*) se ha hecho una detección deficiente. Ambas presentan una característica común: poseen un nivel medio de gris muy elevado. Más concretamente, tan sólo un 15% de los píxeles de dichas mamografías se encuentran por debajo del umbral definido, cuando el valor medio de las restantes se sitúa en un 24%. Por tanto, nuestro programa no responde de manera óptima ante imágenes anormalmente oscuras.

Otra de las peculiaridades que se aprecia al analizar las imágenes resultado es que las grandes calcificaciones son descartadas por nuestro programa. Esto se debe a dos causas: la primera de ella radica en que dichas calcificaciones son visibles para cualquier especialista y su detección por computador resulta innecesaria; la segunda implica al tiempo de ejecución. En una de las versiones de prueba que hicimos conseguimos localizarlas, pero para ello debíamos reducir considerablemente el tamaño del bloque analizado y, por tanto, aumentar el número de bloques con la consecuente ralentización del proceso.

En el siguiente apartado, vamos a ver dos mamografías en concreto: la que mejor resultados ha obtenido, la que peor.

6.3 EJEMPLOS.

6.3.1 Mamografía con una detección perfecta: c16_93rc.

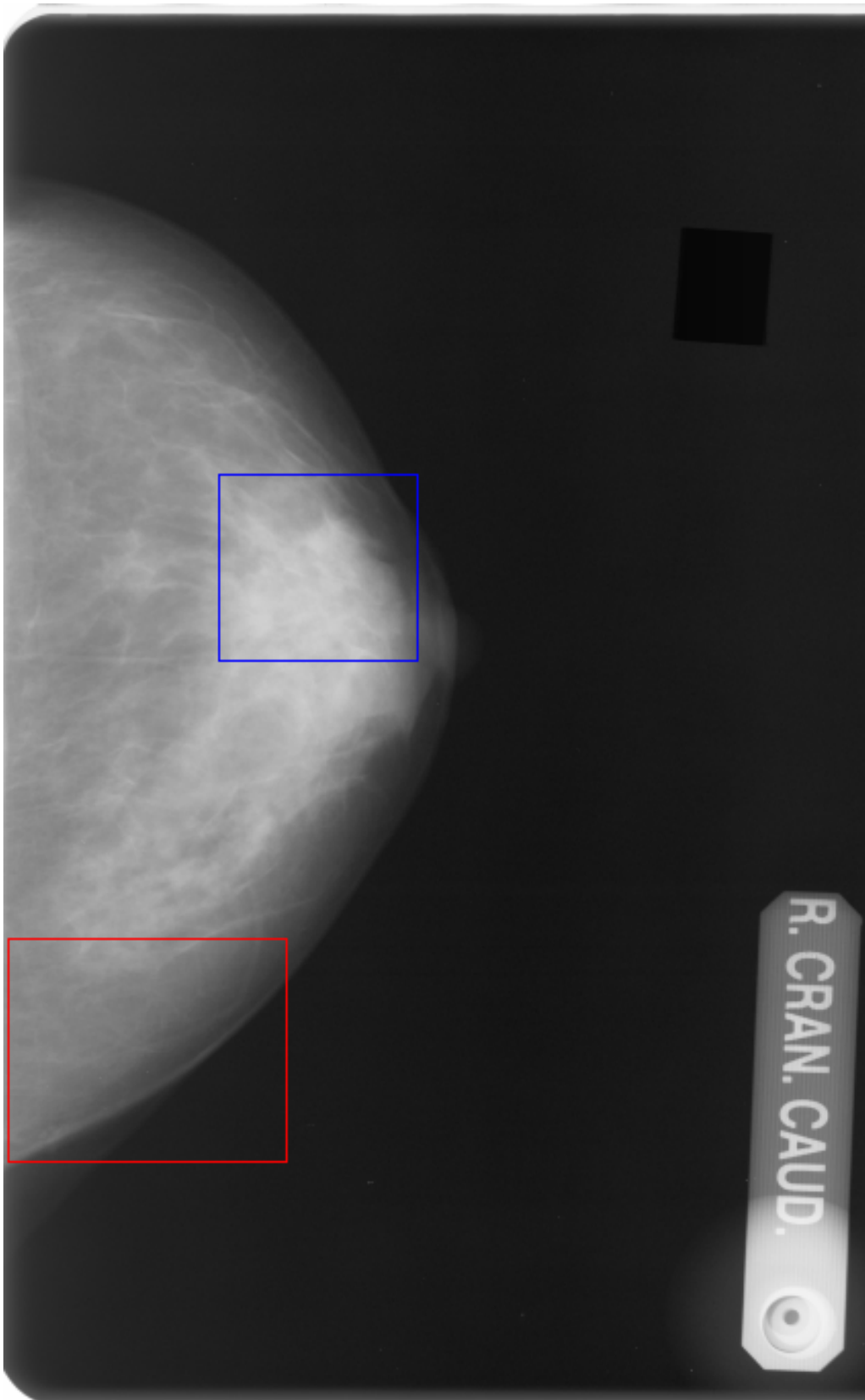


Figura 6.2. Mamografía c16_93rc en baja resolución.

En la imagen podemos apreciar que hemos creado dos zonas. La **zona azul** se corresponde con una sección donde predominan los verdaderos negativos o, dicho de otra forma, donde existen numerosos candidatos que nuestra SVM se encargará de eliminar. Veamos esta zona ampliada para tener un análisis pormenorizado de dicha zona.

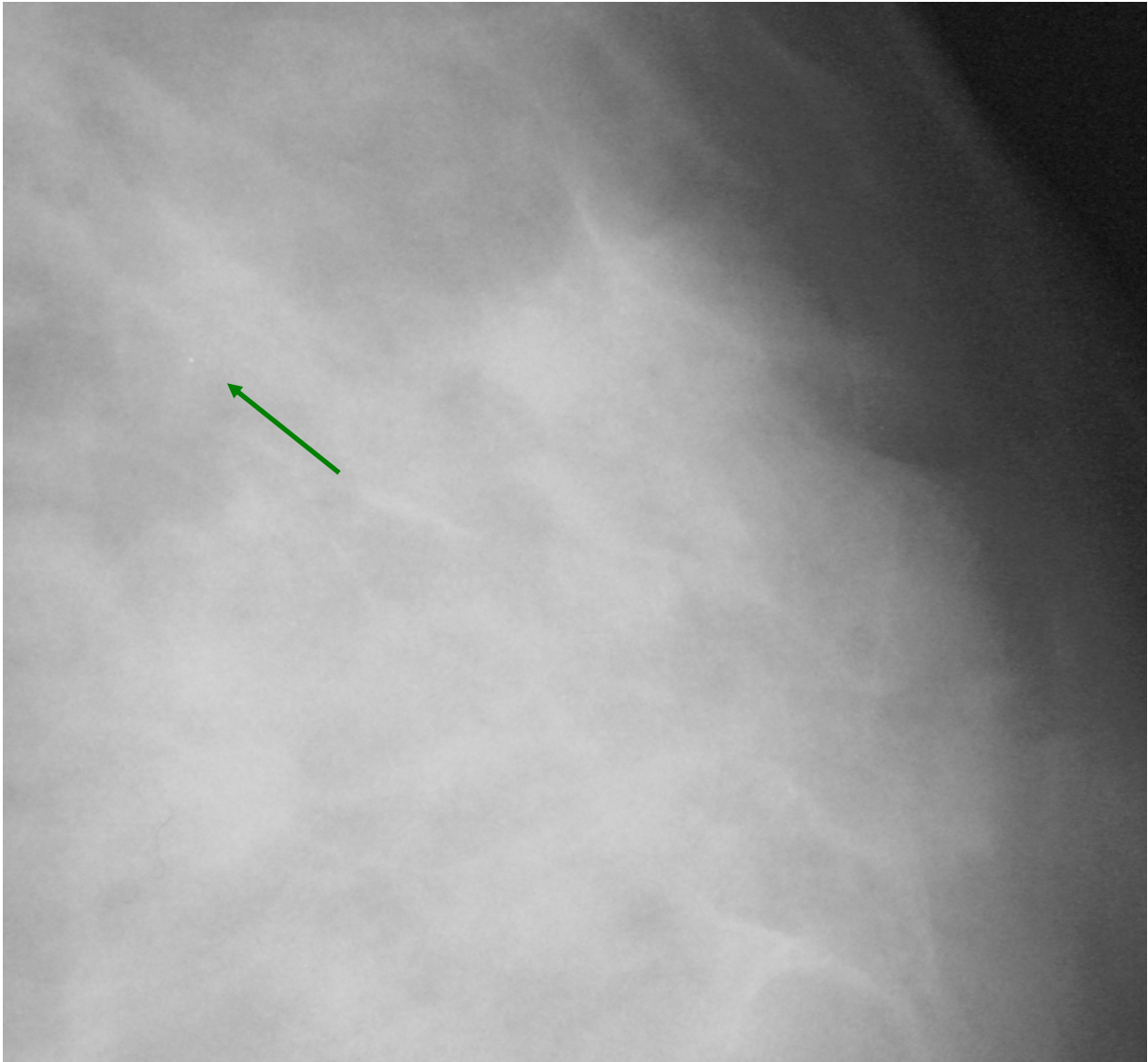


Figura 6.3. Detalle de la zona azul de la mamografía c16_93rc.

Hemos generado una flecha en verde, con un programa de edición de imágenes, con la finalidad de señalar la única microcalcificación existente en la región que estamos estudiando. Se puede considerar que dicha región tiene un valor medio de gris bajo, circunstancia en la que nuestro algoritmo ha respondido con bastante fiabilidad.

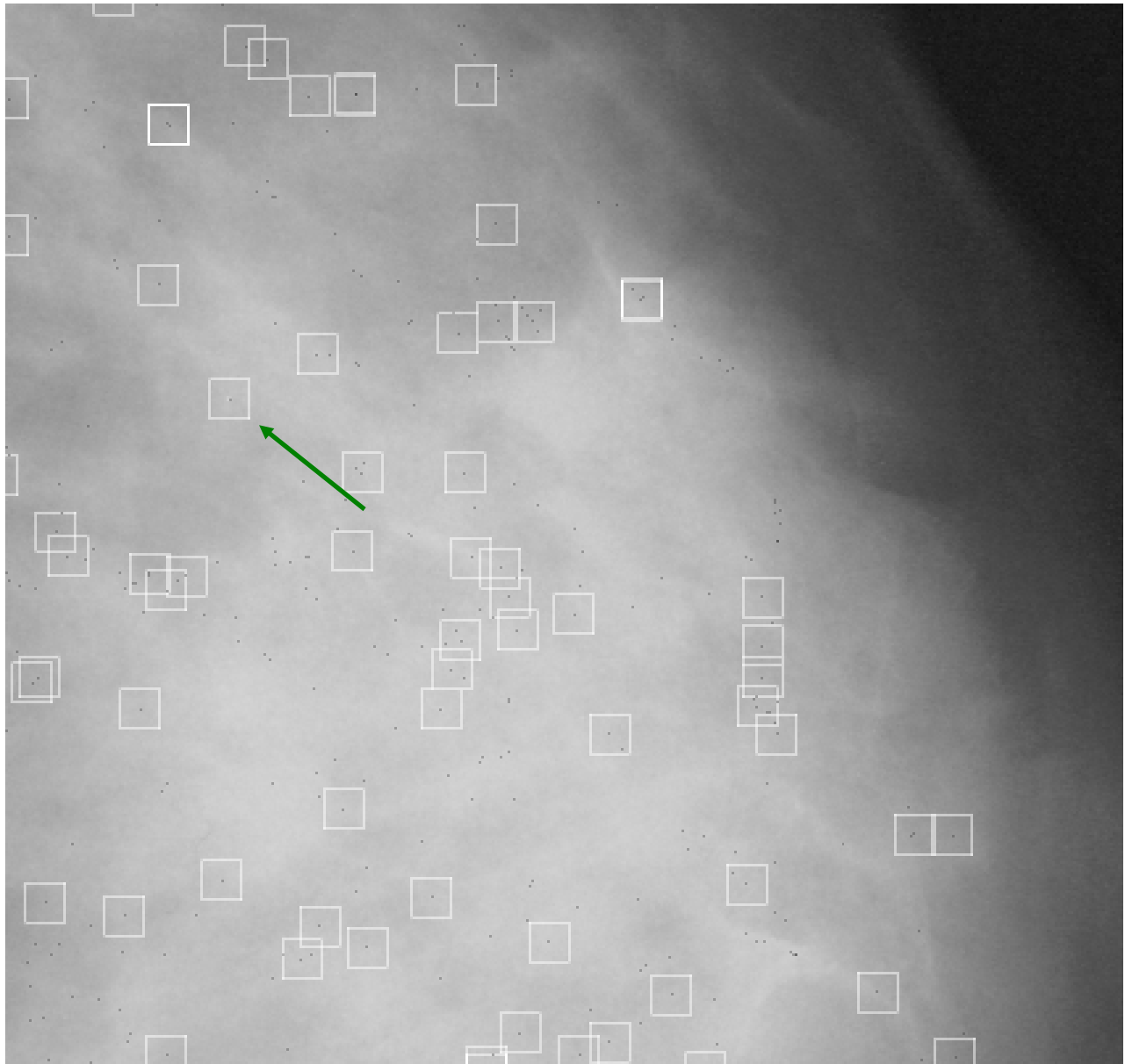


Figura 6.4. Zona azul de la mamografía c16_93rc con los pre-candidatos (puntos negros) y candidatos (cuadrados blancos) señalados.

Observamos aquí que nuestro algoritmo de predicción lineal ha hallado multitud de precandidatos así como la calcificación real (señalada por la flecha verde). Podría pensarse que tan excesivo número es resultado de una deficiente predicción, pero esto no es así. Si nos fijamos bien, alrededor de cada pre-candidato hay una región levemente más clara. Por tanto, dicho algoritmo ha cumplido a la perfección con sus expectativas ya que será la SVM quien decida según su morfología y según su tamaño si esos candidatos son o no considerados como microcalcificación. Mostramos, finalmente, la imagen obtenida tras aplicar la SVM:

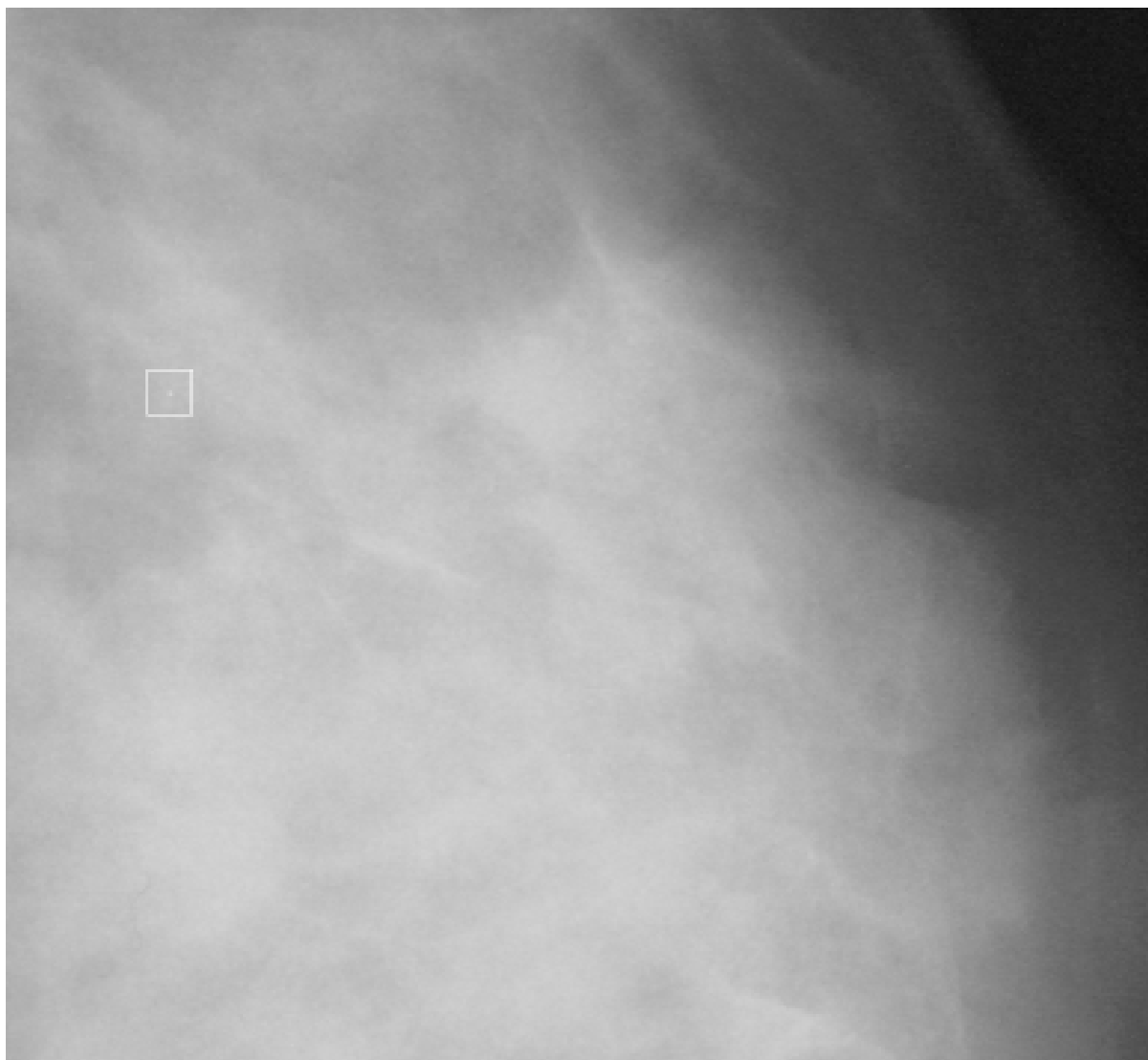


Figura 6.5. Microcalcificación bien detectada en la mamografía c16_93rc.

Para completar el nuestro estudio, vamos a ver qué resultados se obtuvieron en la **zona roja** de la mamografía. Ésta se corresponde con una región de valor medio de gris alto pero con microcalcificaciones bien marcadas. La calidad de los candidatos obtenidos en esta zona es excelente, ya que como se comprobará, casi ninguno de ellos es descartado.

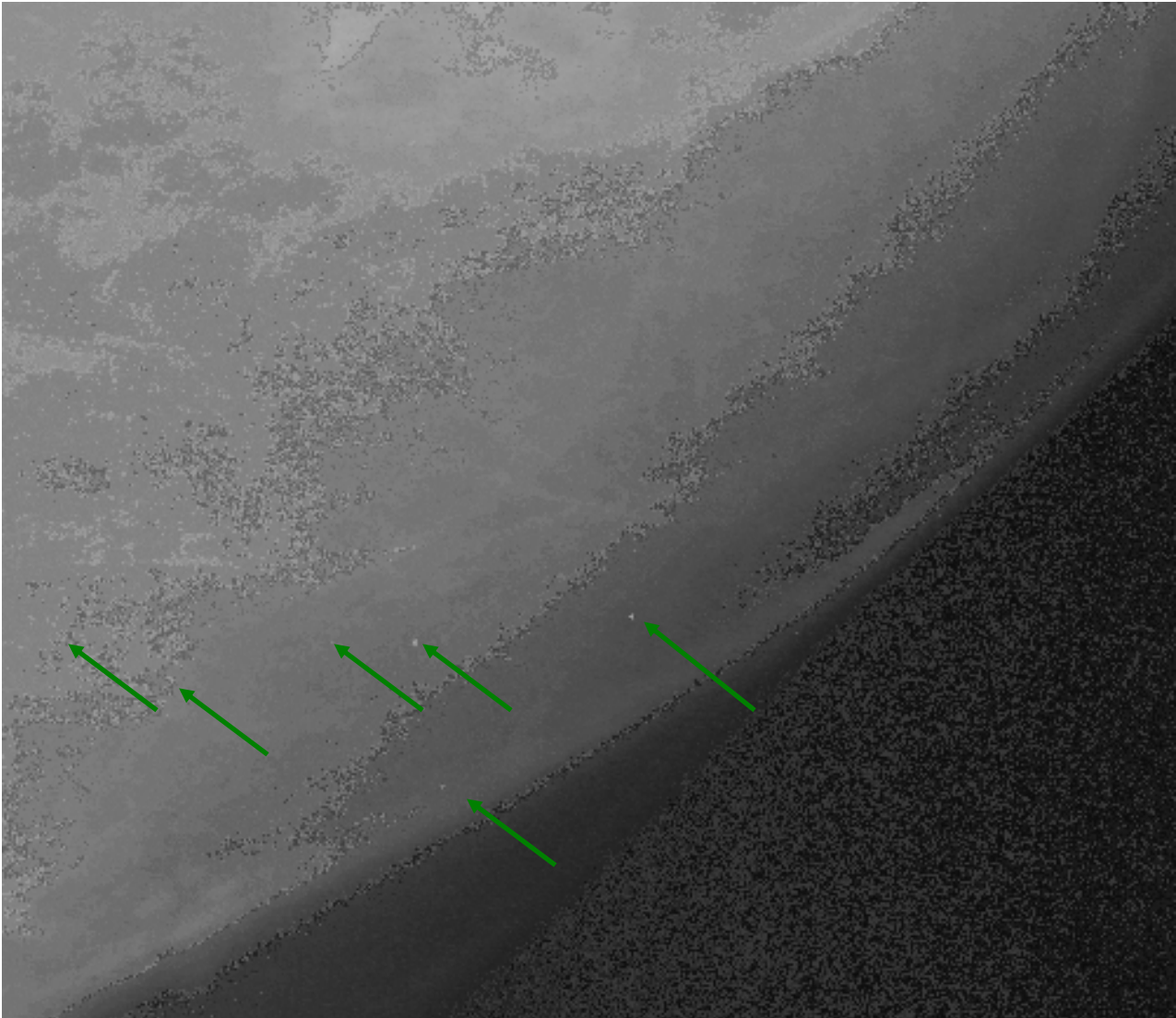


Figura 6.6 Ampliación del cuadrado rojo de la mamografía c16_93rc.

Señalamos con flechas verdes aquellas microcalcificaciones que son detectables a simple vista. No obstante, y aquí está uno de los principales objetivos de esta herramienta, tras la ejecución de nuestra aplicación, observaremos que varias de ellas se nos habían pasado por alto.

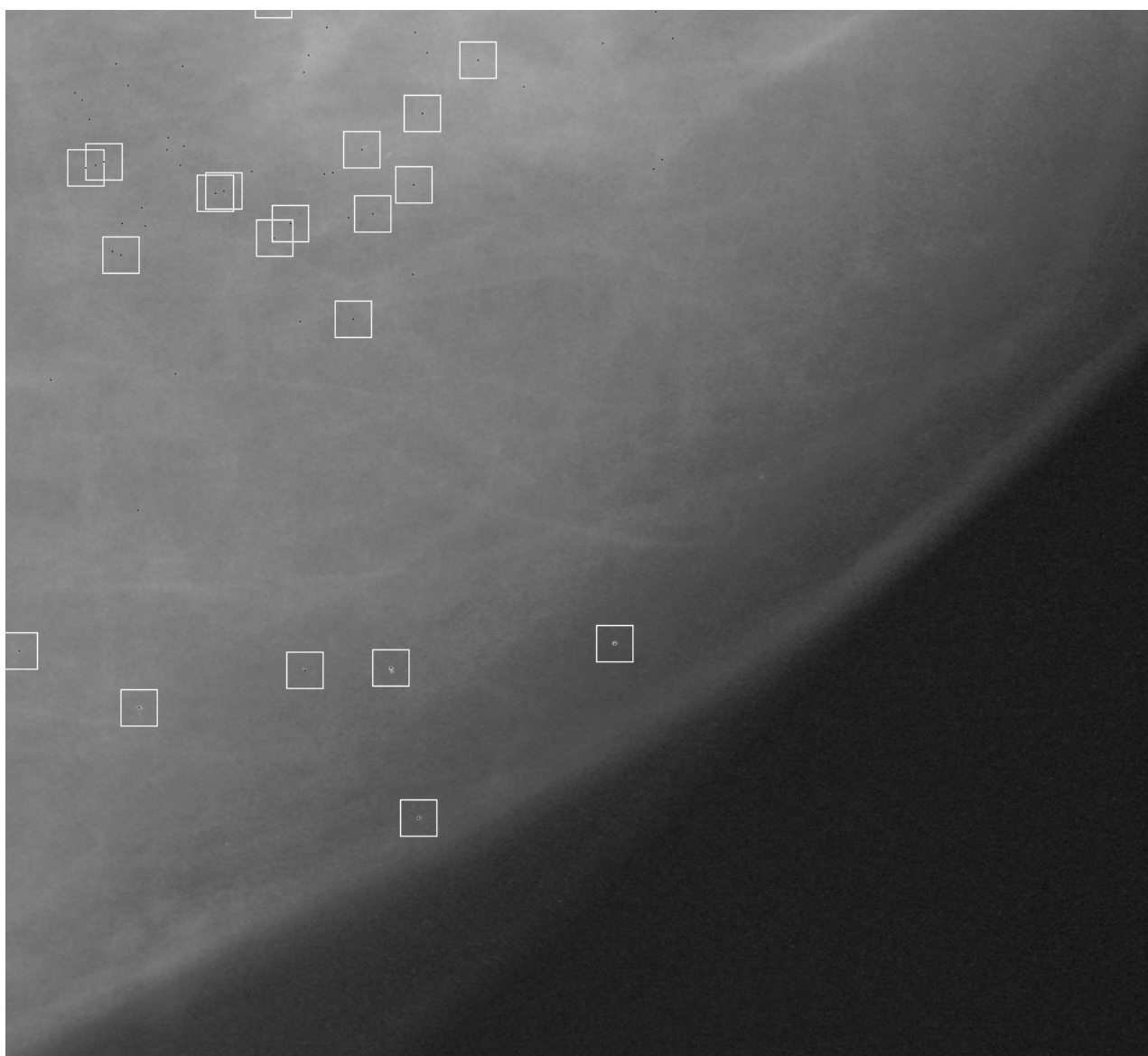


Figura 6.7. Zona azul de la mamografía c16_93rc con los pre-candidatos (puntos negros) y candidatos (cuadrados blancos) señalados.

Finalmente, mostramos las microcalcificaciones obtenidas.

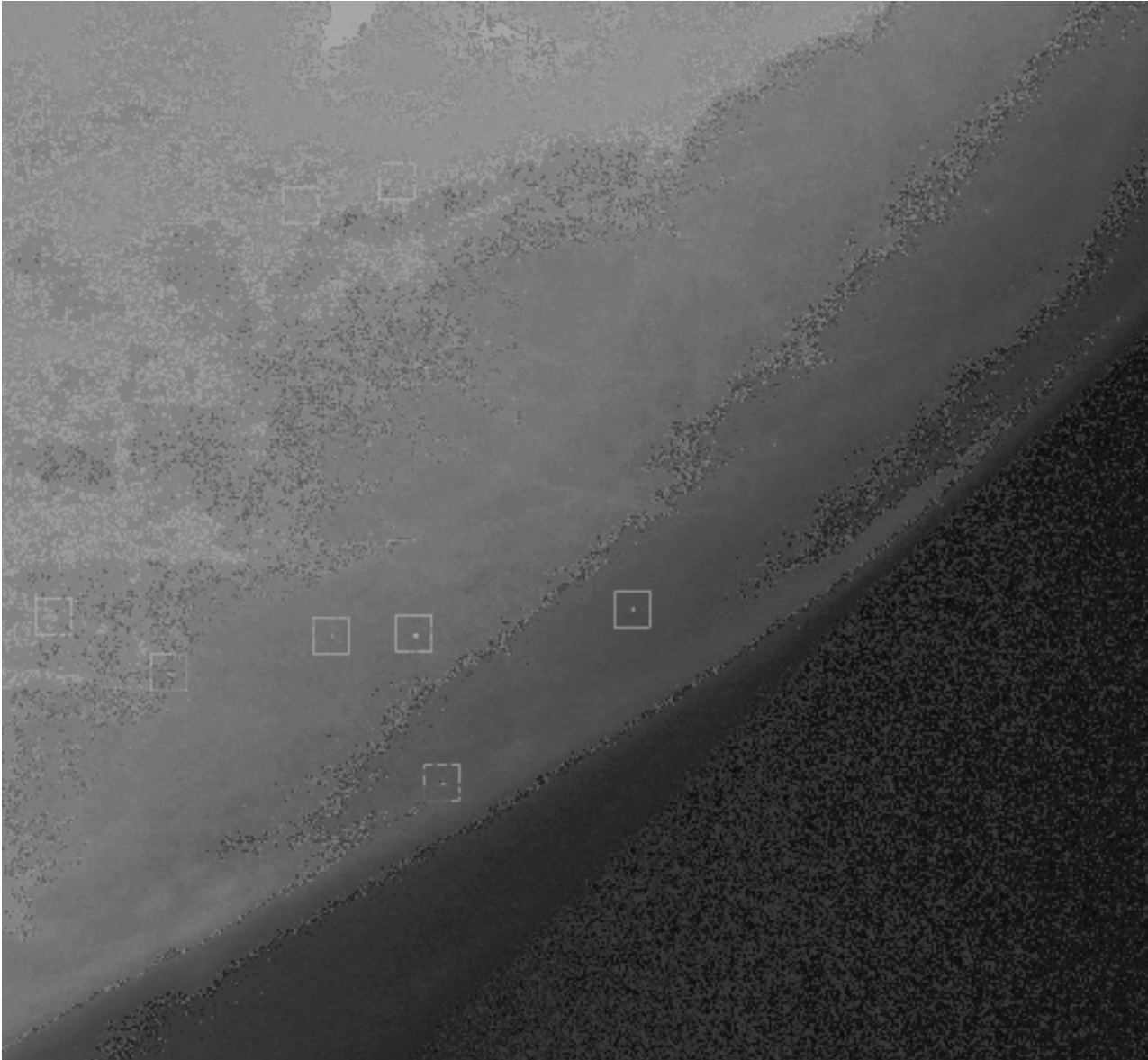


Figura 6.8 Microcalcificaciones bien detectadas en la mamografía c16_93rc.

Para concluir, mostramos una tabla resumen con los valores de los parámetros obtenidos:

C16_93RC	CANDIDATOS	MICROS	VP	FP	VN	FN	SENS.	PPV	ESPEC.	NPV
AHORA	444	52	52	0	392	0	100%	100%	100%	100%
ANTES	722	52	48	4	664	6	88.9%	92.3%	99.4%	99.1%

6.3.2 Mamografía con una detección deficiente: c5_96ro.

Vamos a examinar ahora una mamografía en la que ha habido muchos falsos negativos. Debemos aclarar, no obstante, que este resultado ha sido localizado en sólo dos mamografías de nuestra base de datos, lo que supone aproximadamente un 5%.

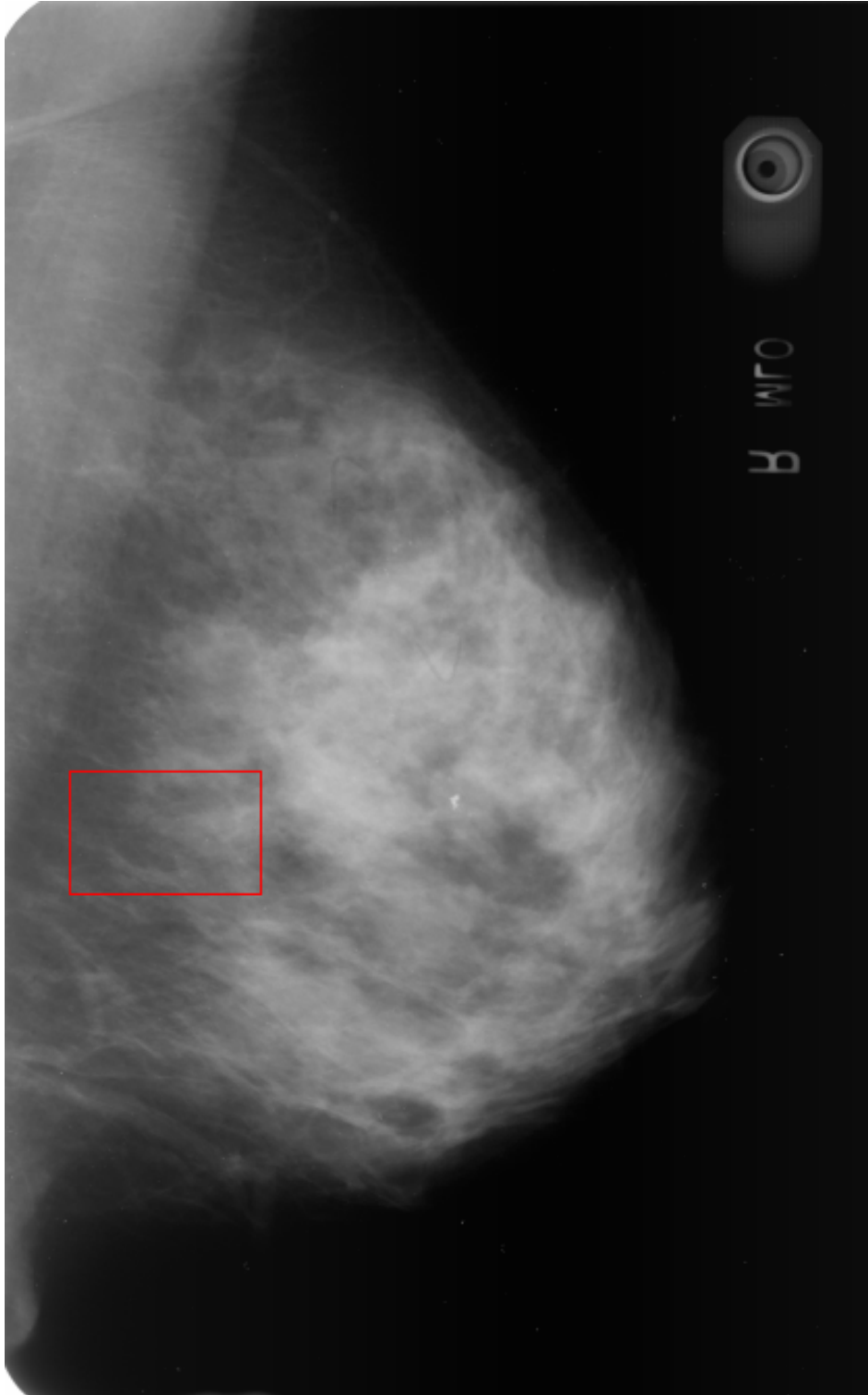


Figura 6.9. Mamografía c5_96ro en baja resolución.

Dado que la imagen expuesta anteriormente está en una resolución reducida, la granularidad introducida hace que muchas de las microcalcificaciones no sean vistas. Por esa razón, mostramos, a continuación, la zona enmarcada en rojo ampliada.



Figura 6.10. Recuadro rojo de la mamografía c5_96ro ampliado.

Vemos que es una zona con una alta densidad de microcalcificaciones y que, además, se trata de una mamografía oscura, hecho que como ya hemos comentado, hará que nuestro algoritmo responda peor. Para empezar, nuestro algoritmo de predicción ya ha pasado por alto tres microcalcificaciones.

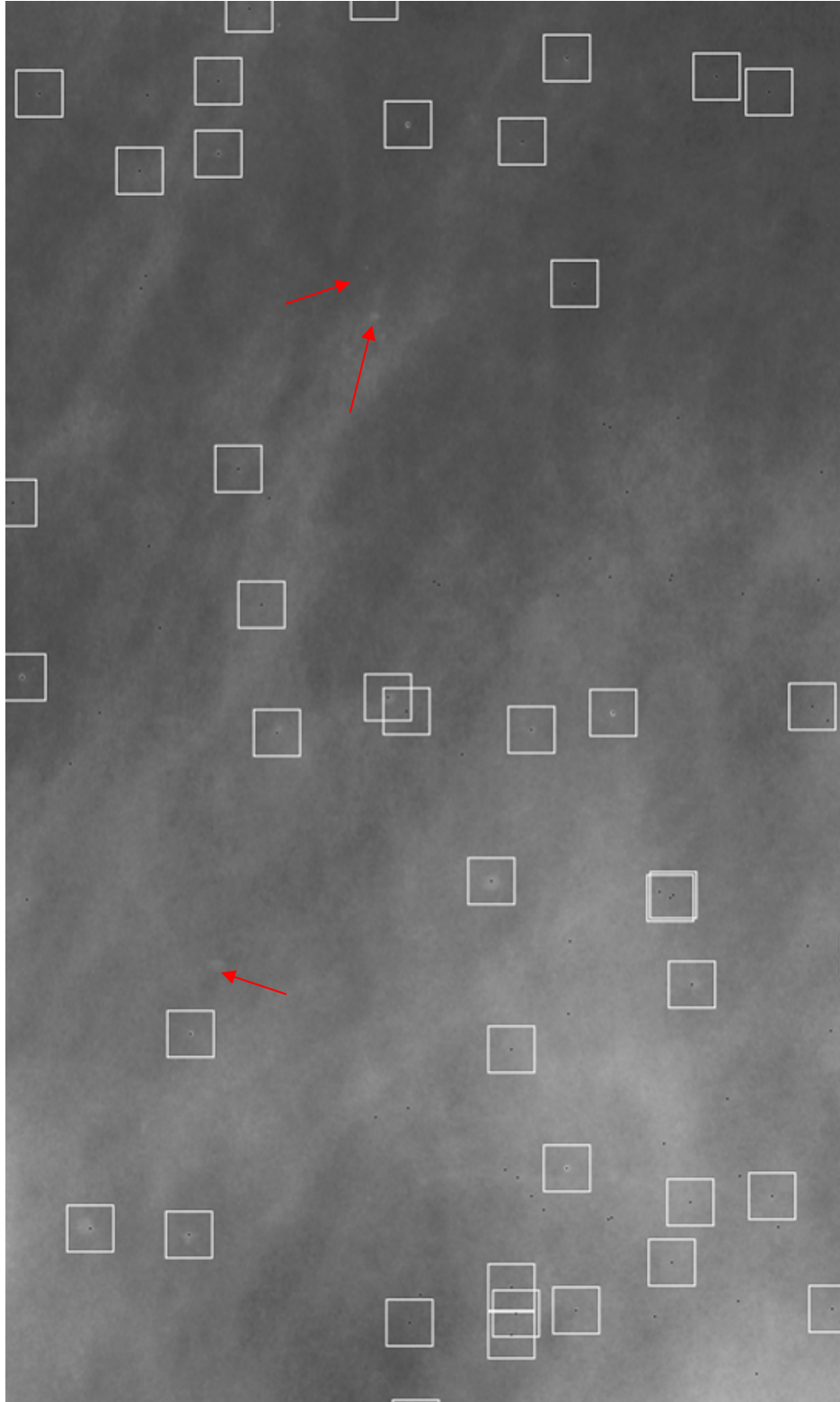


Figura 6.11. Obtención de candidatos en la mamografía c5_96rc.

Vemos, finalmente, la imagen resultado. A las tres FN anteriores, la SVM suma otros tres más:

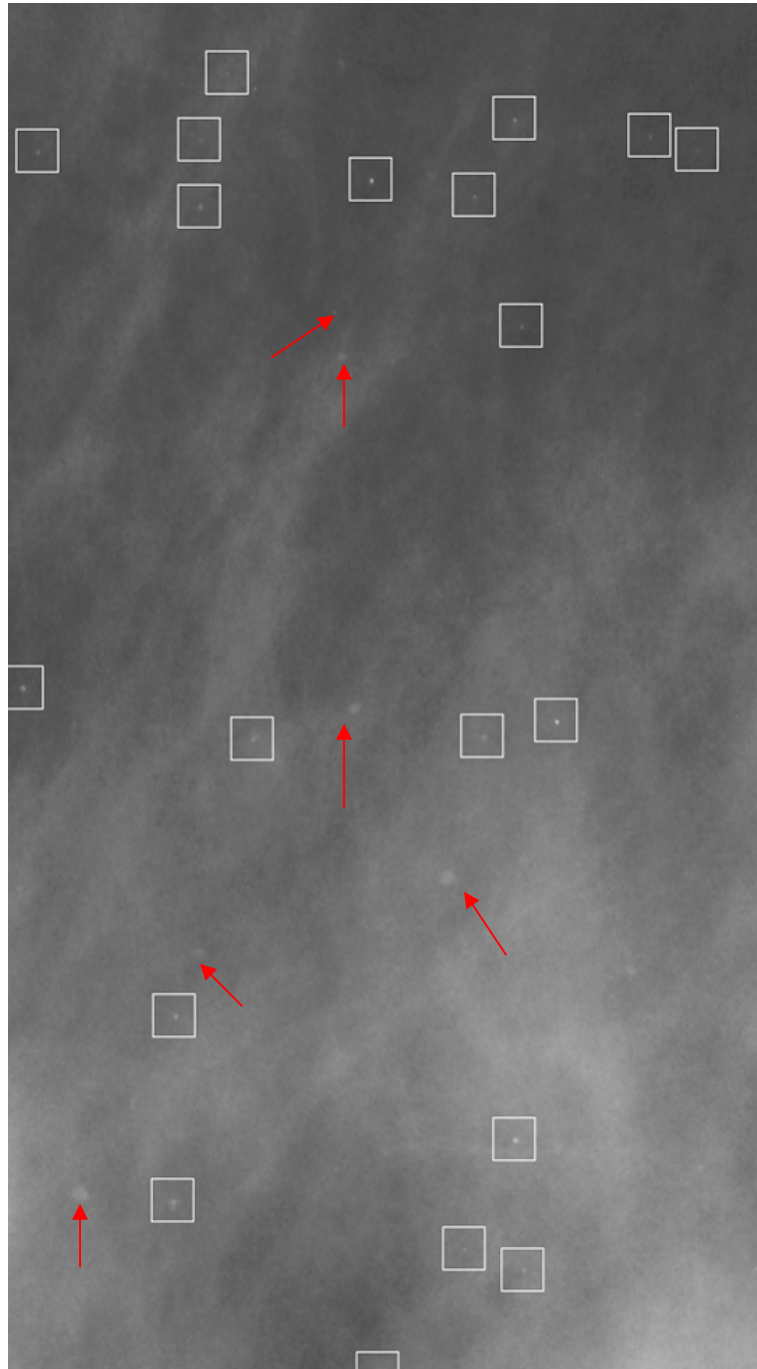


Figura 6.12. Microcalcificaciones bien detectadas en la mamografía c5_96ro.

Mostramos una tabla resumen con los valores de los parámetros obtenidos en la que apreciamos que, a pesar de ser considerado este ejemplo como el de una detección deficiente, la sensibilidad es superior al 94%:

C5_96RO	CANDIDATOS	MICROS	VP	FP	VN	FN	SENS.	PPV	ESPEC.	NPV
AHORA	1514	366	365	1	1125	23	94.1%	99.7%	99.9%	98.0%
ANTES	1712	366	321	4	1343	44	87.9%	98.7%	99.7%	96.8%

6.4 Conclusiones.

Ha llegado el momento de recapitular y analizar la validez de nuestra herramienta. Partíamos de una situación realmente difícil de mejorar, ya que los resultados que se habían obtenido (sensibilidad superior al 90%) eran muy buenos con respecto a otros métodos propuestos, como por ejemplo aquellos que están basados en el empleo de la *transformada wavelet*. Sin embargo, tras una ardua tarea de optimización de código, hemos conseguido no sólo mejorar los resultados sino, además, conseguir un bajísimo coste computacional. Más concretamente, hemos rozado la perfección en la mayoría de los casos (sensibilidad media del 96%) y sólo en dos de ellos hemos obtenido resultados que podríamos considerar como malos (sensibilidad del 88%). En lo que más hemos avanzado ha sido en el valor predictivo positivo o, lo que es lo mismo, cuando algoritmo dice que hay una microcalcificación lo dice con una fiabilidad del 99% frente al 89% que teníamos anteriormente. Si hay algo que no se le puede negar al resultado final obtenido es la extraordinaria sensibilidad que ha demostrado tener, lo que nos ha permitido ver determinadas microcalcificaciones que, a simple vista, eran imperceptibles por el ojo humano.

Disponíamos de una extensa colección de mamografías, lo que nos ha permitido aplicar nuestro algoritmo a todas las situaciones imaginables. Así, nos hemos enfrentado a mamografías muy claras, donde el contraste entre la microcalcificación y su entorno era mínimo; a otras muy oscuras, donde el valor medio de gris de la imagen hacía que el umbral habitual no fuera óptimo e incluso a algunas que ocupaban un porcentaje muy pequeño de la imagen completa. Cada una de ellas aportaba alguna peculiaridad que hacía mejorar día a día nuestro algoritmo. Y así, hemos conseguido tener los resultados tan buenos como robustos, ya que es de esperar que ante nuevas mamografías nuestro programa siga respondiendo con porcentajes de acierto similares.

No debemos olvidar que esta aplicación surge con la idea de complementar la opinión de un experto y jamás con la perspectiva de suplantarla. Debemos resaltar que todos los diagnósticos realizados han sido contrastados por el Dr. Nicolás Mohedano Mohedano, especialista en oncología del hospital *Ciudad de Jaén*. Del mismo modo, la Dra. M^a Reyes Ortiz Vallejo ha puesto a nuestro servicio todos sus conocimientos sobre la materia, guiándonos en todo momento durante la elaboración del mismo.

Las mamografías de las que disponemos son de una alta resolución, hecho que conlleva el manejo de ficheros de más de 50 MB en la mayoría de los casos y por tanto habrá que ser especialmente cuidadosos a la hora de programar si no queremos que nuestro tiempo de ejecución se dispare. Cuando nos enfrentamos por primera vez al código del programa, nos encontramos con una estructura, de alguna manera, caótica. Muchas habían sido las personas que habían trabajado con anterioridad sobre este código y poco había sido el tiempo dedicado a realizar una programación estructurada. Se ha mejorado extraordinariamente en cuanto a lo que a modularidad se refiere, hecho que resultará de extrema utilidad si en un futuro se decide seguir mejorando. Asimismo, hemos eliminado multitud de bucles fútiles, consiguiendo reducir su tiempo de ejecución a un 33%.

En cuanto al algoritmo utilizado en la determinación de píxeles candidatos, ha sido demostrado sobradamente su eficiencia y, por tanto, descartamos la idea de buscar otro. El único pero que se le podría poner es el elevado número de precandidatos que obtiene. No obstante, el tener una herramienta tan potente como novedosa, como es una máquina de vector soportes (SVM), hace que la importancia de este hecho sea mínima ya que su casi perfecta calibración hace que obtener unos resultados óptimos.

Quizá, el único inconveniente que le podemos encontrar es que necesitamos dos entornos de trabajo distintos para completar un análisis completo. En un principio se pensó en unificar ambos entornos en uno sólo pero esto implicaba perder eficiencia, ya que de sobra es conocido por todos la eficiencia que muestra MatLAB para el tratamiento de imágenes y la potencia de C para la gestión de memoria y la gestión de E/S. Si se deseara distribuir esta aplicación o simplemente generalizar su uso en la ayuda al diagnóstico, se deberá crear un entorno de trabajo mucho más cómodo para el profesional, aunque esto conlleve una pérdida de eficiencia computacional, ya que ahora mismo se requiere tener conocimientos de MatLAB y acceder a la línea de comandos del sistema operativo. Por tanto, convertir todo este código a un lenguaje *visual*, sería una tarea del todo recomendable para crear un entorno más agradable al usuario.