



PONTIFICIA
UNIVERSIDAD
CATÓLICA DE
VALPARAÍSO



Sebastián Antonio Vergara Inostroza

Predicción de interrupciones en fusión nuclear mediante detección de anomalías

Informe Proyecto de Título de Ingeniero Civil Electrónico



**Escuela de Ingeniería Eléctrica
Facultad de Ingeniería**

Valparaíso, 04 de marzo de 2019



Predicción de disrupciones en fusión nuclear mediante detección de anomalías

Sebastián Antonio Vergara Inostroza

Informe Final para optar al título de Ingeniero Civil Electrónico,
aprobada por la comisión de la
Escuela de Ingeniería Eléctrica de la
Facultad de Ingeniería de la
Pontificia Universidad Católica de Valparaíso
conformada por

Sr. Gonzalo Alberto Farías Castro
Profesor Guía

Sr. Gabriel Enrique Hermosilla Vigneau
Segundo Revisor

Sr. Sebastián Fingerhuth Massmann
Secretario Académico

Valparaíso, 04 de marzo de 2019

Dedicado a mis padres Jenny y Marco.

Agradecimientos

En mis años universitarios y para la elaboración de este trabajo de titulación quiero agradecer a mis padres Jenny y Marco, los cuales fueron un apoyo incondicional, dándome ánimo, cariño y contención día a día. A mi pareja Camila, que me acompañó en este proceso y me dio las fuerzas para continuar cuando más lo necesitaba, mis amigos Enrique, Miguel y Osvaldo, que fueron parte fundamental desde el primer día, apoyándonos mutuamente y superando todas las dificultades juntos. Finalmente, agradecer a mis profesores que me guiaron, motivaron y que me entregaron la formación necesaria para llevar a término mi carrera y lograr ser un profesional.

Valparaíso, 4 de enero de 2019

S.V.

Resumen

La Fusión Nuclear aplicada a la generación de energía es uno de las metas más importantes a nivel global, ya que de esta manera es posible cambiar el paradigma respecto al escenario energético mundial. A pesar de la amplia gama de beneficios y grandes avances que tiene esta tecnología, ésta presenta el inconveniente de las interrupciones, las cuales terminan de manera súbita el desarrollo normal del plasma y que genera en definitiva gastos tanto temporales como económicos.

Para mitigar o evitar los fenómenos disruptivos es necesario detectar precursores que anticipen estos sucesos, de acuerdo a esto se han implementado un gran número de algoritmos, típicamente basado en aprendizaje automático. En este trabajo, se presenta un sistema de predicción de interrupciones basado en detección anomalías (DPAD), desarrollado a través de una base de datos provenientes del dispositivo tokamak JET y en el cual se alcanzó en el mejor caso una tasas de verdaderos positivos de 96.25(%) (77/80) y falsos positivos de 5.8(%) (24/414). Adicionalmente, se destaca del sistema propuesto su sencillez y que puede ser ajustado para detectar otros fenómenos interesantes que sucedan en el proceso de fusión.

Esta tesis está estructurada en siete capítulos. Los capítulos 1 y 2 introducen a la fusión nuclear y la problemática a abordar, dónde se destacan los principales conceptos, la importancia de esta tecnología y un estudio sobre el estado del arte. En el capítulo 3 se estudia el enfoque de detección de anomalías, definiendo los tipos que existen y una clasificación de los métodos aplicados a este enfoque. En el capítulo 4 se presenta el predictor propuesto a través de las diferentes etapas y criterios que lo componen, mostrando ejemplos ilustrativos y las expresiones matemáticas que los rigen. A continuación, los capítulos 5 y 6 se muestra detalladamente el proceso de entrenamiento mediante algoritmos genéticos y los resultados obtenidos. Al final de este capítulo se analizan los resultados respecto a las señales más utilizadas por el sistema y el porcentaje de disparo disruptivo generados por criterio, para finalmente desarrollar una comparativa con otros predictores, destacando las cualidades positivas y negativas del modelo propuesto. Por último, en el apéndice de esta tesis se muestra la implantación del predictor propuesto entrenado mediante el método DPFS (entrenamiento desde cero), el cual busca simular la puesta en marcha de un nuevo dispositivo de fusión.

Palabras claves: fusión nuclear, escenario energético mundial, interrupción, aprendizaje automático, detección de anomalías, algoritmos genéticos, entrenamiento desde cero.

Abstract

Nuclear Fusion applied to power generation is one of the most important goals globally, because in this way it is possible to change the paradigm for the global energy scenario. Despite the wide range of benefits and great advances of this technology, it has the disadvantage of disruptions, which suddenly end of the normal development of plasma and ultimately generates both temporary and economic costs.

To mitigate or avoid disruptive phenomena it is necessary to detect precursors that anticipate these events, according to this a large number of algorithms have been implemented, typically based on automatic learning. In this work, a system of prediction of disruptions based on anomalies detection (DPAD) is presented, developed through a database from the tokamak JET device and in which it reached in the best case a true positive rates of 96.25(%) (77/80) and false positives of 5.8(%) (24/414). Additionally, the simplicity of the proposed system stands out and can be adjusted to detect other interesting phenomena that occur in the fusion process.

This thesis is structured in seven chapters. Chapters 1 and 2 introduce nuclear fusion and the problem to be addressed, which highlights the main concepts, the importance of this technology and a study on the state of the art. In chapter 3 the anomaly detection approach is studied, defining the types that exist and a classification of the methods applied to this approach. In chapter 4 the proposed predictor is presented through the different stages and criteria that compose it, showing illustrative examples and the mathematical expressions that govern them. Next, chapters 5 and 6 show in detail the training process using genetic algorithms and the results obtained. At the end of this chapter we analyze the results regarding the most used signals most used by the system and the percentage of disruptive tripping generated by criteria, to finally develop a comparison with other predictors, highlighting the positive and negative qualities of the proposed model. Finally, the appendix of this thesis shows the implementation of the proposed predictor trained by the DPFS method (training from scratch), which seeks to simulate the implementation of a new fusion device.

Key words: nuclear fusion, global energy scenario, disruption, machine learning, anomaly detection, genetic algorithms, training from scratch.

Índice general

Introducción.....	1
1 Fusión nuclear.....	4
1.1 Escenario energético mundial.....	4
1.1.1 Energías amigables con el medio ambiente.....	5
1.2 Energía nuclear.....	6
1.2.1 Fisión.....	6
1.2.2 Fusión.....	7
1.3 Dispositivos de fusión nuclear.....	8
1.3.1 Stellarator TJJ-II.....	9
1.3.2 Tokamak JET.....	10
1.4 Futuro de la fusión.....	12
1.4.1 Proyecto ITER.....	12
1.4.2 Proyecto DEMO.....	13
2 Descripción de la problemática y motivación.....	14
2.1 Introducción.....	14
2.2 Disrupción.....	14
2.2.1 Base de datos – Señales.....	16
2.3 Enfoque machine learning.....	18
2.3.1 Estado del arte.....	18
2.3.2 Sistema de caja negra.....	20
2.3.3 Sistema de caja transparente.....	20
2.4 Objetivos del proyecto.....	21
2.4.1 Objetivo general.....	21
2.4.2 Objetivos específicos.....	21
3 Detección de anomalías.....	22
3.1 Introducción.....	22
3.2 Tipos de anomalías.....	23
3.2.1 Anomalías puntuales.....	23
3.2.2 Anomalías colectivas.....	23

3.2.3 Anomalías contextuales	23
3.3 Etiqueta de los datos	24
3.4 Métodos de detección	24
3.4.1 Clasificación.....	25
3.4.2 Vecino más cercano	25
3.4.3 Clustering	25
3.4.4 Estadística	26
3.5 Detección de anomalías para la predicción de interrupciones	26
4 Predictor de Disrupciones basado en Detección de Anomalías (DPAD)	28
4.1 Lag-plot - Mapa autorregresivo	28
4.1.1 Definición.....	28
4.1.2 Patrones lagplot	28
4.1.3 Lag-plot base de datos	30
4.2 Detección de anomalías	32
4.3 Arquitectura predictor	32
4.3.1 Etapa 1	33
4.3.2 Diagrama de bloques etapa 1	36
4.3.3 Etapa 2	36
5 Entrenamiento DPAD.....	40
5.1 Optimización metaheurística	40
5.1.1 Algoritmos genéticos.....	41
5.2 Entrenamiento por conjunto	41
5.3 Implementación etapa 1	42
5.4 Implementación etapa 2	44
5.4.1 Entrenamiento de refinamiento	46
6 Resultados.....	48
6.1 Etapa 1.....	48
6.2 Etapa 2.....	49
6.2.1 Resultados entrenamiento por criterio	49
6.2.2 Resultado entrenamiento de refinamiento.....	50
6.3 APODIS	52
6.4 Comparativa.....	53
6.5 Análisis DPAD.....	54
6.5.1 Curva ROC.....	57
Discusión y conclusiones.....	63
Bibliografía	66
A Entrenamiento desde cero	70
A.1 Predictores desde cero	70

A.1.1 Implementación método DPFS	70
A.1.2 Veen Predictor	71
A.2 Resultados entrenamiento desde cero sistema DPAD	72
A.3 Comparativa.....	74

Introducción

Los avances tecnológicos ocurridos en las últimas décadas a nivel mundial, han originado un desarrollo significativo tanto económico como social, generando a su vez una transformación respecto al modo en que nos relacionamos, pensamos y actuamos. Es bien sabido que el principal consumo energético de los usuarios es del tipo eléctrico, esto debido a que la gran mayoría de las tecnologías basan su funcionamiento en este tipo de energía, provocando en definitiva que ésta tenga un crecimiento explosivo y sostenido que es necesario suministrar. Los datos que otorga el banco mundial indican que desde el año 1971 hasta el 2014 el consumo de energía eléctrica per cápita a nivel mundial ha crecido desde 1199.37(*kWh*) hasta 3127.36(*kWh*) y sigue creciendo a un ritmo acelerado. Adicionalmente, este efecto se ve incrementado si se considera el aumento abrupto en el tamaño de la población mundial.

De manera complementaria al alza en el consumo energético, existen otros dos tópicos de alta relevancia en la actualidad, la escasez de combustible y el impacto al medio ambiente, los cuales han sido poco tomados en cuenta en el siglo pasado, pero que hoy en día han tomado vital importancia para la sociedad. Finalmente, todo recae en la pregunta: ¿Cómo podemos suministrar toda esta energía sin afectar el medio ambiente y a partir de combustible abundantes en mundo?

Esta pregunta deriva a una amplia gama de complejos retos para la humanidad, que necesariamente deben ser abordados. Es así como surgió la opción de la generación de energía a través de fusión nuclear o también conocida como la energía de las estrellas. En el año 1926, Arthur Eddington [1] propuso que existe un límite donde la presión de la energía que emiten las estrellas sobre la atmósfera es lo suficientemente grande como para equilibrar su propia fuerza gravitacional. Este científico británico describió a las estrellas como esferas de gas a alta temperatura, compuestas principalmente por hidrógeno y donde la presión de las fuerzas gravitacionales puede alcanzar una temperatura que convierte los gases en plasma (el cuarto estado de la materia). La temperatura y la presión en las estrellas fusionan los núcleos de hidrógeno en el plasma produciendo helio y energía. La fusión nuclear como energía renovable busca generar electricidad reproduciendo la reacción que ocurre en las estrellas y que conceptualmente es capaz de atacar directamente las problemáticas del escenario energético mundial.

Según lo expuesto han surgido grandes programas internacionales que han desarrollado dispositivos con la finalidad de experimentar y estudiar a fondo este fenómeno. Existen dos enfoques para estos dispositivos, los de confinamiento inercial y los de confinamiento magnéticos, siendo estos últimos los mayormente estudiados.

Uno de los proyectos que destacan a nivel mundial, es el construido bajo el auspicio de la Unión Europea. JET (Joint European Torus) es un dispositivo experimental de fusión nuclear de tipo Tokamak, que fue construido a partir del año 1978 e inició sus operaciones 1984, desde esta fecha hasta 1991 logró demostrar que es posible mantener confinado el plasma mientras se produce la fusión nuclear. Gracias a los resultados obtenidos a partir de JET, la ambición de llevar a cabo esta forma de generación de energía se ha ampliado, por lo que se han ideado proyectos como ITER (International Thermonuclear Experimental Reactor), el cual busca consolidar la generación de energía en base a fusión nuclear mirado del punto de vista de la rentabilidad y seguridad. Este proyecto es financiado por cerca de 35 naciones, donde destacan China, Japón, Estados Unidos, India, Corea del Sur y la Unión Europea. ITER no busca generar energía, por esto y como siguiente etapa está el proyecto DEMO, el cual tiene como objetivo construir la primera planta de generación de energía eléctrica mediante la fusión nuclear por confinamiento magnético.

Adicionalmente a la gran cantidad de dificultades que involucra la construcción y funcionamiento inicial en los dispositivos Tokamak como JET e ITER, está el alto grado de inestabilidad del proceso y la complejidad en la dinámica del plasma (magnetohidrodinámica), siendo esta última altamente no lineal y por consecuencia no ha posible aún de ser totalmente controlado y dando origen en definitivas a los fenómenos repentinos conocidos como disrupciones.

Las disrupciones son eventos donde el plasma confinado presenta una inestabilidad súbita, lo cual conlleva a daños en las paredes de los dispositivos y a su vez en pérdidas económicas y temporales, incluso son capaces de generar daños irreversibles. Existen diferentes métodos relacionados con la mitigación y evitación de las disrupciones, pero es necesario anticiparse a ellas. Por consiguiente, es fundamental elaborar sistemas que detecten y determinen cuando sucederán estos eventos, de manera que el proceso sea suficientemente estable y seguro. Los fenómenos físicos que conducen a estos eventos son muy complejos y no lineales, por lo tanto, se ha abordado esta problemática a través de sistemas de aprendizaje automático, pues éstos aprenden sobre los datos y, por ende, no se requiere determinar un modelo explícito del comportamiento de las señales que se muestrean en cada experimento. ITER establece que los sistemas predictores de disrupciones requieren un mínimo de acierto de 95(%) y las falsas alarmas que no superen un 5(%), además respecto a los tiempo de anticipación se necesitan generar alarmas que entreguen como mínimo un plazo de 10(*ms*) para actuar [2]. Es así como se utilizados una amplia gama de enfoques para la predicción de disrupciones con mayor y menor éxito, de las cuales destacan las basadas en aprendizaje automático como Redes Neuronales Artificiales (RNA), Máquina de Vectores Soporte (SVM), Lógica difusa (FZ), Redes Neuronales Recurrentes (RRN) y Detección de Anomalías (AD).

El último enfoque mencionado es el abordado en este trabajo, y el cual alude a una divergencia, incompatibilidad, discordancia o desvío de una norma o de una práctica, en este caso en particular las normas serían extraídas de aquellas señales donde el plasma no tiene un comportamiento disruptivo. DPAD, acrónimo de Disruptions Predictor based Anomaly Detection (Predictor de Disrupciones basado en Detección de Anomalías en español), tiene como objetivo no solo mejorar los niveles de predicción y tiempos de anticipación, sino que además el sistema busca ser sencillo y que los parámetros que intervengan en él tengan un sentido lógico, es decir, que no se comporte como una caja negra.

Por otra parte, el sistema desarrollado en este trabajo no solo está enfocado a la detección de precursores disruptivos, sino que para ser implementado en la determinación de cualquier fenómeno de interés que sucede en el proceso de fusión. Esto último es fundamental debido al hecho de que los dispositivos de fusión generan una cantidad de datos que es imposible analizar de manera manual.

1 Fusión nuclear

1.1 Escenario energético mundial

La estimación de la Organización de las Naciones Unidas (ONU) respecto al crecimiento de la población mundial desde el 1990 hasta el año 2017, indica un aumento de más 2000 millones de personas en el mundo (desde 5300 millones a 7550 millones) y sus predicciones muestran que para el año 2030 la población alcanzaría los 8551 millones [3].

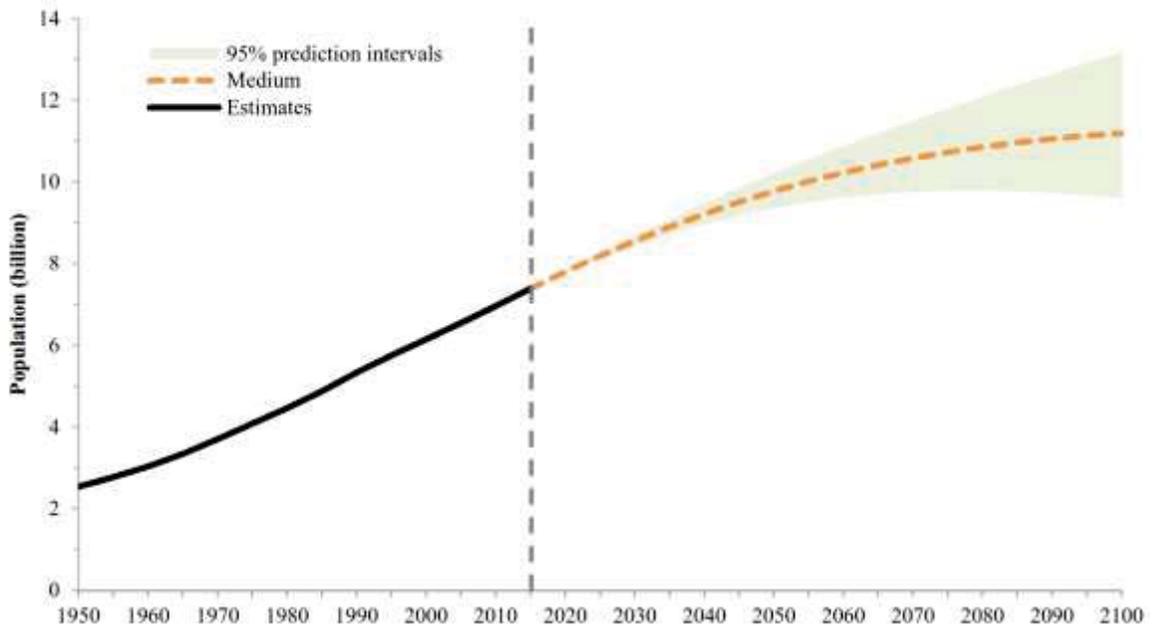


Figura 1-1: Desarrollo y estimación de la Población Mundial [3]

Es intuitivo aseverar que este crecimiento explosivo en la población conlleva un aumento en la demanda energética mundial, lo cual es correcto, de acuerdo a la Agencia Internacional de la Energía en su publicación World Energy Outlook 2018 se estima que para el año 2040 la demanda mundial tendrá un incremento entorno al 30(%)

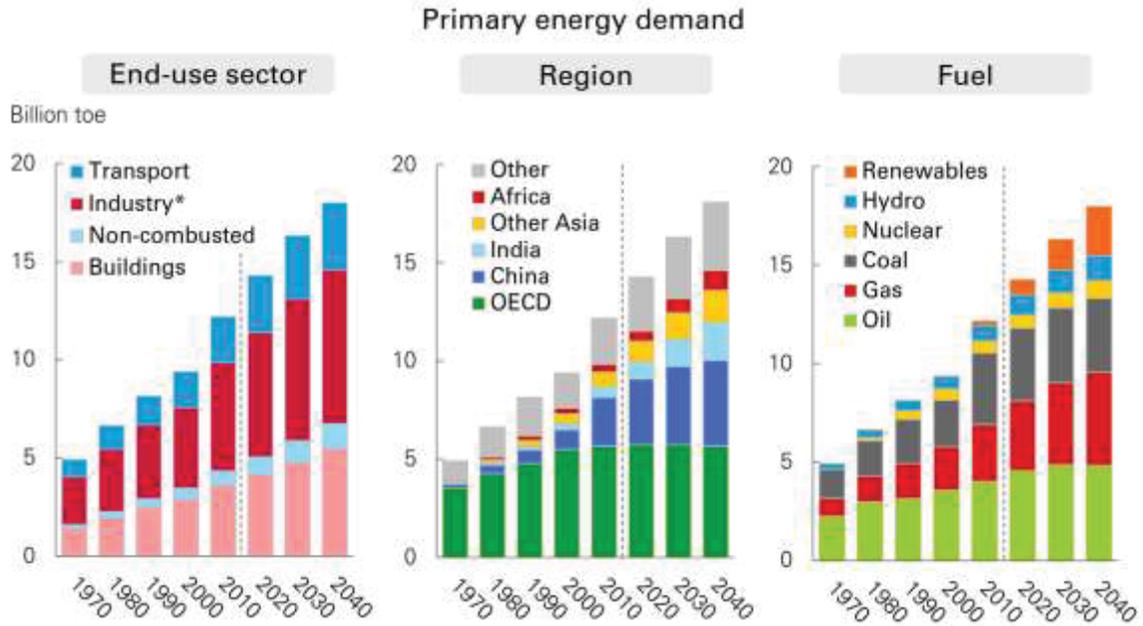


Figura 1-2: Desarrollo y estimación de la Demanda Energética [4]

Por otro lado, y en consecuencia de lo anterior, las reservas energéticas mundiales convencionales han tenido una disminución considerable debido principalmente al hecho de que este tipo de energías no renovable poseen costos de extracción más económicos. En la Tabla 1-1 se muestra el nivel de reserva para las principales fuentes energéticas mundiales según el estudio que se presenta en [5].

Tabla 1-1: Reserva mundial de los combustibles fósiles [5]

	Reserva (<i>Mtep</i>)		Producción (<i>Mtep</i>)		Duración (años)
	2011	1993	2011	1993	
Carbón	891530	1031610	7520	4474	> 100
Petróleo	223454	140676	3973	3179	56
Gas Natural	209742	141335	3518	2176	55

Por todo lo anterior planteado se hace de gran importancia evaluar nuevas maneras de generación de energía, de acuerdo a esto se destacan principalmente dos; la energía nuclear y las energías renovables. Estas últimas, aunque han tenido un auge en los últimos años debido fundamentalmente por ser amistosas con el medio ambiente, tan solo corresponden un pequeño porcentaje de la demanda energética y los escenarios futuros, aunque mejoran, siguen siendo insuficientes, lo anterior se debe a la poca rentabilidad y baja eficiencia que poseen.

1.1.1 Energías amigables con el medio ambiente

Pese a que en los últimos años la humanidad ha tenido grandes avances en tecnología, salud, ciencia, industria y otras áreas, el consumo de combustibles fósiles continúa siendo la principal fuente de energía. Es así como se han liberado grandes cantidades de CO₂ y otros gases de efecto

invernadero, causando lo que se conoce como “Calentamiento Global” o “Cambio climático”. Pero, ¿Qué consecuencias tiene el calentamiento global en la humanidad? Mirado desde el punto de vista particular, existen probables locaciones donde este fenómeno sea beneficioso, en cambio para la visión globalizada problemas como el calor extremo, desastres naturales, distribución de las infecciones, sequías, entre otros, pueden generar grandes problemáticas mundiales. Según la Organización Mundial de la Salud (OMS) se estima que entre los años 2030 y 2050 el cambio climático será causante de 250.000 defunciones anualmente [6].

A modo de resumen, los tres factores que intervienen más profundamente con el escenario energético mundial son el aumento constante de la demanda energética, reserva de combustibles fósiles en decadencia y la preocupación respecto al cambio climática global. Cada una de estos es importante abordarlos y buscar métodos de generación de energía que los ataquen directamente. Es así como la fusión nuclear surge como respuesta al contexto energético mundial.

1.2 Energía nuclear

El enfoque Nuclear entrega una posible solución a las problemáticas planteadas, esto debido a que no sólo es capaz de generar grandes niveles de energía, sino que además los combustibles utilizados poseen grandes reservas a nivel mundial. Ahora bien, este método se compone en dos sub-áreas, la Fisión y Fusión Nuclear.

1.2.1 Fisión

La división de un núcleo atómico es el proceso denominado como Fisión Nuclear, básicamente este proceso en la práctica se genera a partir de la colisión de núcleos de Uranio (elemento con alto número atómico), del cual nacen dos nuevos núcleos más pequeños e inestables y además se libera una gran cantidad de energía, tal energía es transformada en calor para finalmente hacer la conversión a electricidad. Este proceso genera una reacción en cadena que debe ser controlada, hoy en día no existen mayores problemas en la seguridad y el control del proceso, no obstante, la utilización de elementos con peso atómico elevado genera residuos radioactivos nocivos para el hombre y adicionalmente, elementos como el Uranio tienen un límite de reserva establecido.

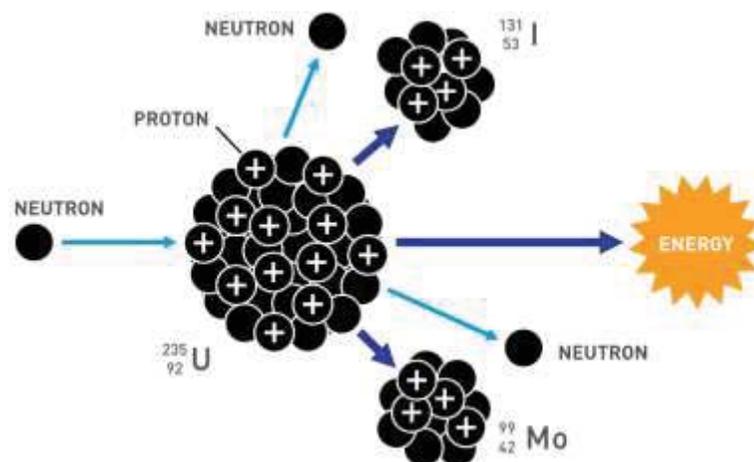


Figura 1-3: Esquema ilustrativo del proceso de Fisión (fuente: <https://cna.ca>)

1.2.2 Fusión

La Fusión Nuclear está presente constantemente en nuestras vidas a $149.6M(km)$ de distancia, en el sol. Es ahí donde altos niveles de presión y temperatura generan que átomos ligeros como el Hidrogeno (H), sufran un proceso denominado Fusión, formando así un nuevo elemento más pesado como el Helio (He) y liberando grandes sumas de energía. El gran desafío es traer este proceso a la tierra, en la práctica la fusión se provoca a partir de los isótopos de Hidrogeno como el Deuterio y Tritio, los cuales conforman un nuevo átomo de Helio, acompañado de la liberación de un Neutrón y Energía (radiación), pero las dificultades están en imitar las condiciones de alta temperatura del Sol. En la Figura 1-4 se muestra de forma gráfica el proceso descrito.

Para entender la razón de esta liberación de energía, basta con realizar el equilibrio de masas, en donde la masa antes y después del proceso de fusión son diferentes (aunque manteniendo el número de partículas, es decir que el neutrón junto con el Helio sumados poseen menos masa que el Deuterio y el Tritio al inicio, esta diferencia de masa se transforma en energía de acuerdo a la mundialmente famosa ecuación de Albert Einstein, que relaciona la Energía (E) con el producto de la masa (m) y la velocidad de la luz (c) (1-1).

$$E = mc^2 \quad (1-1)$$

Las ventajas que posee la utilización de este proceso para la generación de energía son: capacidad de manejar grandes niveles de potencia, los combustibles utilizados (Deuterio y Tritio) son en la práctica inagotables y además no genera casi ningún tipo de residuo nocivo para los humanos (a diferencia de la Fisión Nuclear), por el lado de las desventajas se tiene: la gran inestabilidad que posee este proceso y que aún no se logra que la potencia generada sea superior a la utilizada como entrada.

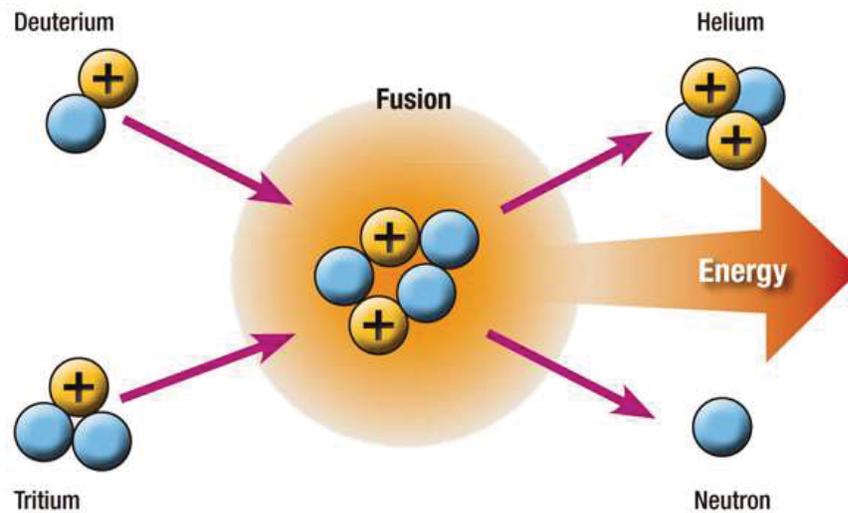


Figura 1-4: Esquema ilustrativo del proceso de Fusión Deuterio-Tritio (fuente: <https://energia-nuclear.net>)

¿Por qué deuterio y tritio?

Para generar la reacción se requieren niveles de temperatura elevados, la probabilidad de que este fenómeno ocurra se ve reflejada a través de lo que se denomina como sección eficaz (cross section en inglés) de la reacción. En la Figura 1-5 el par Deuterio (D) y Tritio (T) muestra un alto nivel de sección eficaz para temperaturas o energías bajas en comparación a otras parejas. Este aspecto, acompañado del hecho de que sus reservas a nivel mundial son abundantes, llevan a optar por ellos.

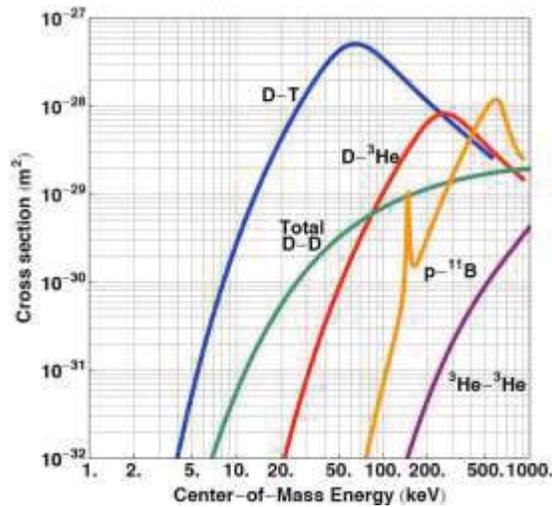


Figura 1-5: Cross section (sección eficaz) para las reacciones más frecuentes en fusión (fuente: <https://physics.stackexchange.com>)

1.3 Dispositivos de fusión nuclear

Como fue mencionado en el apartado anterior, lo que se busca es generar el escenario que se presenta en el sol (alta temperatura y presión), para esto actualmente las investigaciones se basan en dos principales enfoques: confinamiento inercial y magnético. En el confinamiento inercial, una esfera de hidrógeno es comprimida por un láser a temperaturas y densidades tan altas que el criterio de Lawson [7] se cumple, el cual define que se genera el proceso de ignición cuando la multiplicación de la densidad de partículas del plasma η , el tiempo de confinamiento τ y la temperatura T alcanzan cierto valor. Para la fusión de deuterio-tritio, este valor alcanza aproximadamente lo que muestra en la expresión (1-2).

$$\eta\tau T \geq 5 \cdot 10^{21} \left(\frac{sKeV}{m^3} \right) \quad (1-2)$$

Los métodos de confinamiento inercial [8] [9], típicamente se basan en una serie de fases (Figura 1-6). Primero la esfera de hidrogeno se calienta mediante una fuente externa (láseres), esto genera que las capas externas de la esfera se desprendan y se comprima el núcleo logrando una reducción del volumen de 1/1000 veces respecto al original. Finalmente, el pulso externo emite una onda para comprimir el centro de la esfera, esto produce altas temperaturas y presiones suficiente para la ignición, las condiciones anteriores se mantienen por periodo de tiempo muy reducidos (nanosegundos).

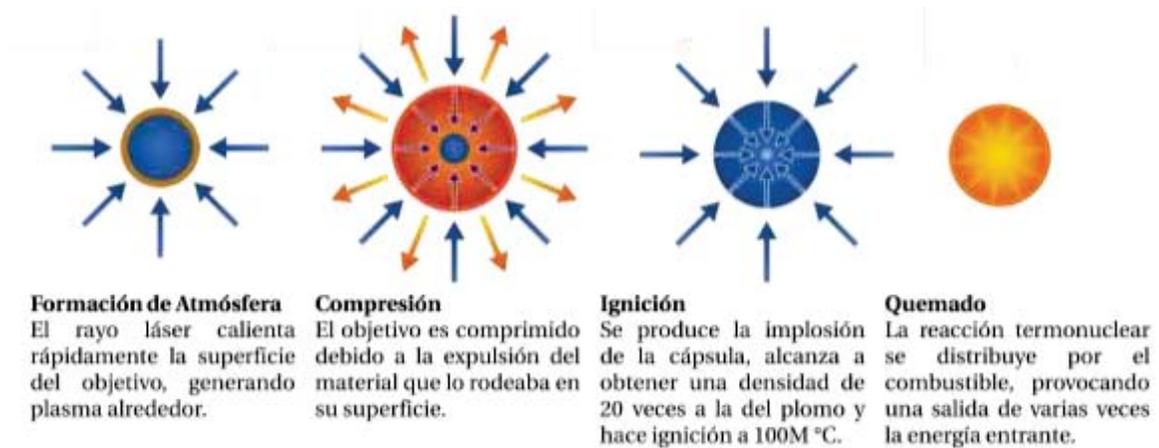


Figura 1-6: Procedimiento del confinamiento inercial (fuente: <https://wikipedia.org>)

Por otro lado, el confinamiento magnético se basa a partir de la fuerza de Lorentz, en donde partículas cargadas realizan un movimiento rotatorio o helicoidal cuando están ante un campo magnético. La idea entonces es que el gas al ser calentado se desioniza, alcanzando un nuevo estado de la materia conocido como Plasma, cargada positivamente y debido a la alta presión y temperatura, los núcleos en el plasma están muy cercanos y se mueven a altas velocidades, de modo que finalmente colisionan y se produce la Fusión Nuclear. Algunos de los reactores nucleares que utilizan el principio de confinamiento magnético son los que se presentan a continuación.

1.3.1 Stellarator TJJ-II

Stellarator viene de la combinación de “*stella*” y “*generator*”, estelar y generador en español, haciendo alusión al proceso de fusión que se genera en estos Astros. Es un tipo de Reactor Nuclear inventado por el físico teórico Lyman Spitzer en 1950 y construido un año después en la Universidad de Princeton, Estados Unidos. Algunos de los dispositivos son el LHD (Japón), W-7AS (Alemania), HSX (Estados Unidos) y el TJ-II (España). En la Figura 1-7 se ilustran los tres tipos de stellarators existentes.

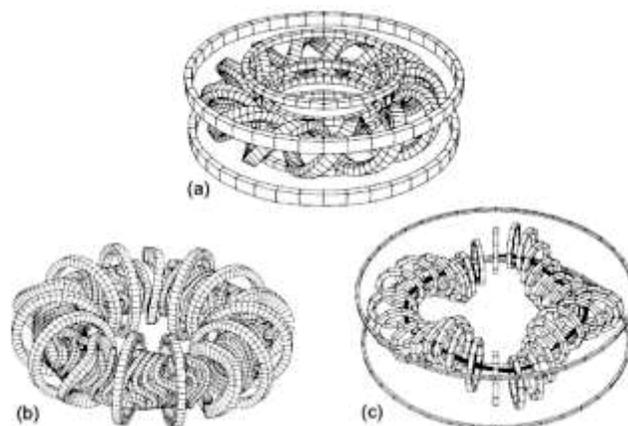


Figura 1-7: Esquemas de los tipos de Stellarators. (a) Torsatrón (desuso). (b) Modular (desuso). (c) Helic

- Torsatrón: configuración con bobinas helicoidales continuas o bobinas separadas que producen campos similares.
- Modular: configuración con un conjunto de bobinas modulares (separadas) y una bobina toroidal torcida [10].
- Heliac: Stellarator de eje helicoidal; configuración donde la complejidad de la geometría envuelve el campo magnético que describe una trayectoria helicoidal.

Por otro lado, TJ-II es un Stellarator tipo heliac flexible [11] (Figura 1-7 (c)) diseñado por CIEMAT (Centro de Investigación Energéticas, Medioambientales y Tecnológicas, España) en colaboración con el ORNL (Laboratorio Nacional Oak Ridge, EEUU) y que fue financiado en parte por la EURATOM (Comunidad Europea de la Energía Atómica), la primera vez que funcionó y produjo plasma fue en el año 1997. Los Stellarator heliac se construyen a partir de bobinas planas circulares distribuidas a lo largo de una hélice enrollada alrededor de una bobina circular. Sus dimensiones son de un radio externo aproximado de $1,5(m)$ e interno de $0,2(m)$, lo que lo transforma en un Stellarator de tamaño medio.

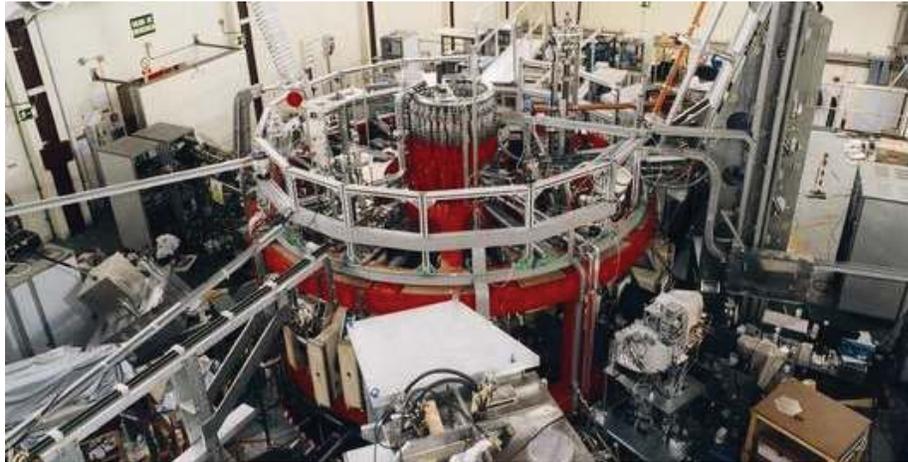


Figura 1-8: Dispositivo Stellarator TJ-II (fuente: <http://www.fusion.ciemat.es>)

1.3.2 Tokamak JET

Tokamak es otro tipo de reactor termonuclear planteado por los físicos soviéticos Ígor Tam y Andréi Sájarov en 1950, su nombre proviene de la unión de las palabras “*toroidalnya kamera*” y “*magnetnaya katushka*” (cámara toroidal y bobina magnética, en español). Se conforma de bobinas encargadas de generar un campo toroidal y por otro lado un campo magnético poloidal que se genera a través de una corriente eléctrica que circula por el mismo plasma a través de inducción, siendo esta una de las principales diferencias con los stellarators. En la Figura 1-9 se muestran los principales elementos que conforman este tipo de reactor.

JET (Toro Común Europeo, sus siglas en español) Figura 1-10, es un reactor del tipo Tokamak situado en las cercanías de Oxford, en el Reino Unido, su construcción se realizó entre los años 1978 y 1982 coordinado por la EURATOM, pero el programa experimental comenzó el 25 de junio de 1983 [12]. Está conformado de acuerdo a lo comentado anteriormente y además posee varios

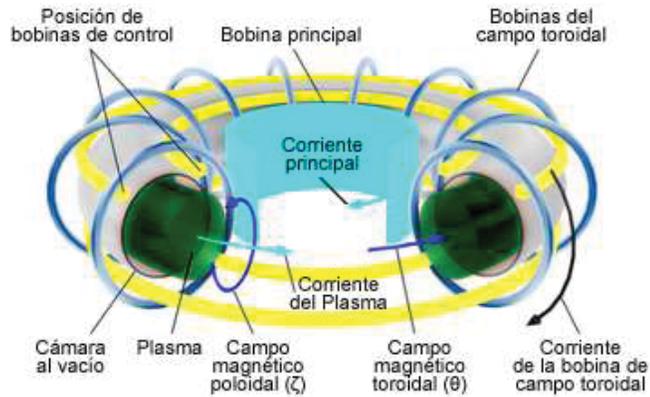


Figura 1-9: Estructura dispositivo Tokamak (fuente: <https://sputnik87.wordpress.com>)

elementos que buscan la estabilización del plasma. El plasma está encerrado dentro del recipiente al vacío en forma de rosquilla que tiene un radio principal de $2.96(m)$ y una sección transversal en forma de "D" de $4.2(m)$ por $2.5(m)$, siendo así uno de los más grandes del mundo en la actualidad.

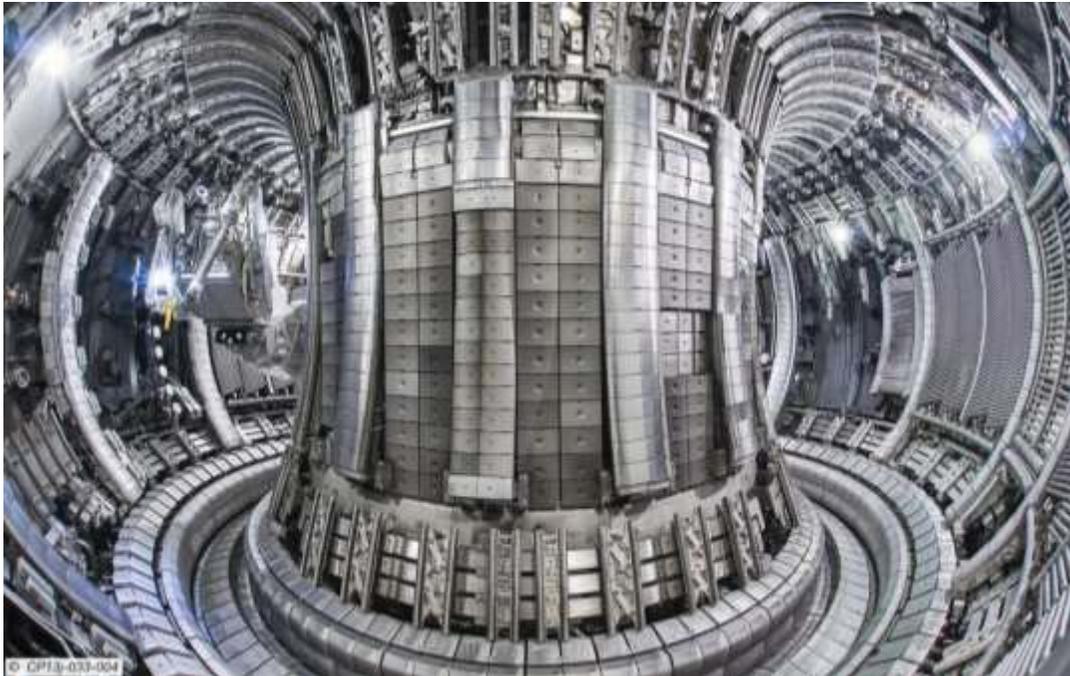


Figura 1-10: Dispositivo Tokamak JET (fuente: <https://www.euro-fusion.org>)

El objetivo esencial de JET es obtener y estudiar un plasma en condiciones y dimensiones que se aproximen a las necesarias en un reactor termonuclear. Estos estudios están orientados a definir los parámetros, el tamaño y las condiciones de trabajo de un reactor Tokamak.

1.4 Futuro de la fusión

Utilizar la Fusión Nuclear como generación de energía es uno de los objetivos más importantes a nivel global, ya que de esta manera habría un cambio de paradigma respecto al escenario energético mundial. Para que este sistema sea viable económicamente, es necesario que genere más energía de la utilizada para realizar el proceso y precisamente esto es lo que busca el proyecto ITER.

1.4.1 Proyecto ITER

Originalmente de sus siglas en inglés International Thermonuclear Experimental Reactor (en español Reactor Termonuclear Experimental Internacional) y también de su significado en latín “camino”, haciendo alusión a que podría conducir a un nuevo sistema de generación de energía, es un proyecto científico a gran escala promovido por 35 naciones como China, India, Japón, Estados Unidos, Corea del Sur y la Unión Europea (este último financia cerca del 45(%) del proyecto) principalmente.

ITER se basa en los reactores termonucleares del tipo Tokamak. Está siendo construido desde el año 2006 y se espera que el primer plasma se genere en diciembre del año 2025 [13] y que tenga duraciones entorno a 30(min). El lugar de construcción es al sur de Francia, en Cadarache, en un sitio de 180 hectáreas (Figura 1-11). Se espera que ITER no funcione continuamente, sino que demuestre la factibilidad del proceso termonuclear para la generación de energía para una potencia estimada de 500(MW), que es equivalente a la capacidad de una planta eléctrica de tamaño mediano. Adicionalmente, la ganancia del proceso sería de $Q = 10$, por lo que la potencia inyectada estaría en valores cercanos a 50(MW).



Figura 1-11: Estructura ITER 2018 (fuente: <https://www.iter.org>)

1.4.2 Proyecto DEMO

DEMO, del acrónimo *DEMONstration Power Station* [14] [15], es un proyecto que va más allá de ITER e intenta demostrar de manera empírica que es posible la generación de energía a través de la fusión nuclear. Su diseño y parámetros finales tienen amplia relación con los resultados que se obtengan de la implementación del proyector ITER, aunque se estima que pueda ser capaz de producir entre 2000(MW) y 4000(MW) para el año 2050.

El valor factor Q esperado para DEMO es mayor o igual a 25, es fundamental que este parámetro sea lo suficientemente grande para que esta forma de generación de energía sea atractivo mirado desde el punto de vista comercial. Según esto, los científicos del área indican que el factor Q está directamente asociado a la cantidad de colisiones que se generan en el proceso, por lo anterior se explica el aumento en la dimensión de los dispositivos (Figura 1-12).

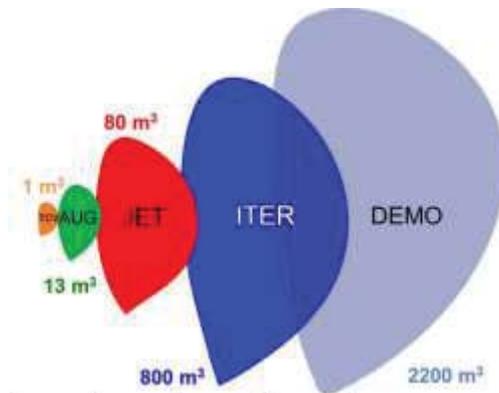


Figura 1-12: Dimensiones dispositivos tokamak [14]

2 Descripción de la problemática y motivación

2.1 Introducción

La Fusión Nuclear como se ha establecido en el capítulo anterior, posee una serie de características que lo posicionan como la mejor opción para la generación de energía del futuro, estas son:

1. capacidad de manejar grandes niveles de potencia,
2. una fuente prácticamente inagotable de combustible,
3. no produce residuos nocivos para el ambiente,
4. la mayor potencia por kilogramo de combustible y
5. técnicamente segura.

A parte de los grandes beneficios que posee este enfoque, existe la desventaja respecto a la alta inestabilidad que sufre el proceso, siendo en concreto la principal problemática el fenómeno denominado como Disrupción Nuclear.

2.2 Disrupción

Las disrupciones son fenómenos que generan la interrupción súbita de los dispositivos nucleares del tipo Tokamak, éstos no han sido posible de ser controlados debido al alto grado de complejidad en la dinámica del plasma, la cual se denominada como magnetohidrodinámica [16] (MHD, dinámica de fluidos conductores de electricidad ante campos eléctricos y magnéticos), que a su vez tiene un comportamiento altamente no lineal. En definitiva y en palabras sencillas, lo que resulta de las disrupciones es un considerable aumento en la inestabilidad, pérdida del confinamiento y finalmente una liberación de energía que puede provocar daños en las paredes del dispositivo, como se observa en las imágenes de la Figura 2-1, lo cual conlleva gastos tanto temporales como monetarios e incluso pueden llegar a generar daños irreversibles.

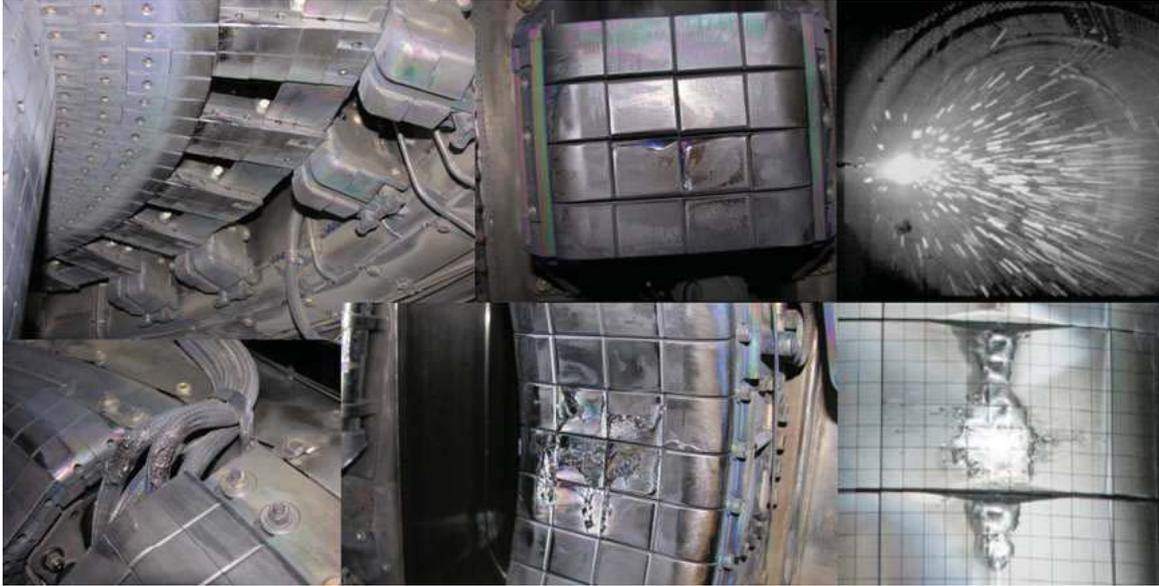


Figura 2-1: Daño por interrupciones en las paredes del JET (fuente: <https://wsx.lanl.gov>)

Gracias a la cantidad de experimentos realizados a través de estos dispositivos, se ha logrado estudiar y establecer límites a diferentes parámetros del proceso para que los fenómenos disruptivos no ocurran. A pesar de lo anterior y en aun cuando el plasma trabaja dentro de los límites operacionales (corriente eléctrica, presión y densidad), es posible que se generen inestabilidades. El momento preciso donde se provoca la interrupción se encuentra bien establecido, debido a que coinciden con variaciones abruptas en las señales extraídas desde plasma, respecto a las fases que suceden en estos fenómenos se pueden distinguir cuatro principalmente:

1. Fase inicial: un cambio en las condiciones del plasma conduce a una configuración menos estable, donde la corriente y la densidad del plasma aumentan ligeramente. Este cambio es causado por factores externos (fallas operativas o mecánicas) o factores internos de plasma (inestabilidades), sin embargo, la variación en las condiciones del plasma es casi imperceptible y prácticamente indetectable.
2. Fase de precursores: los cambios de inicio en la configuración de plasma (fase inicial) alcanzan un punto crítico debido a nuevas inestabilidades MHD detectables y en aumento. La importancia de detectar con suficiente antelación esta fase es crucial. Si los fenómenos se detectan con tiempo suficiente, se pueden implementar técnicas de mitigación para prevenir o reducir los daños. En teoría, estas inestabilidades comienzan como islas magnéticas o modos que giran alrededor del eje del plasma magnético, y crecen exponencialmente en decenas de milisegundos hasta que se genera la interrupción. Un modelo de esta fase fue desarrollado en [17].
3. Thermal Quench (decaimiento térmico): sucede un segundo punto crítico, el perfil de corriente radial se aplanan y la temperatura central desciende drásticamente en unos pocos milisegundos.
4. Current Quench (decaimiento de la corriente): la disminución de la temperatura electrónica de la fase anterior, genera un aumento en la resistividad del plasma, lo que a

su vez provoca una disminución en su corriente y finalmente la energía magnética del campo poloidal se disipa como energía térmica y se pierde el confinamiento. La liberación de esta energía puede ser perjudicial para la integridad del dispositivo y el tiempo de decaimiento depende de las condiciones particulares, la posición del plasma y la sección transversal de éste.

Un ejemplo de una disrupción típica en JET y las fases descritas anteriormente se ilustra en la Figura 2-2.

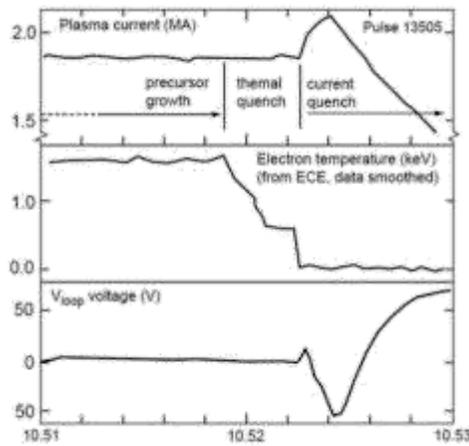


Figura 2-2: Ejemplo de disrupción en JET [17]

Conjuntamente, es relevante destacar que existen diferentes tipos de disrupciones que depende del comportamiento de las fases descritas anteriormente, en [18] se realiza una clasificación para ocho diferentes tipos de interrupciones. Sumado a lo anterior, en [19] se muestra un estudio respecto a las causas de las disrupciones en JET.

A lo largo de los años un gran número de investigadores han buscado la manera de encontrar patrones que indiquen de forma anticipada cuando sucederá una disrupción, de manera de poder evitar o mitigar la problemática [2] [20] [21] [22] [23], es así como se hace atractivo y motivador investigar sobre esta área, principalmente por el margen de mejora que aún existe y debido a la poca cantidad de datos procesados que se extraen de los dispositivos de fusión.

2.2.1 Base de datos – Señales

Gracias al Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas CIEMAT, el cual coopera mediante el otorgamiento de campañas efectuadas en el JET, se consiguió una base de datos correspondiente a las campañas desde la C28 a la C30 (obtenidas entre agosto de 2011 y julio de 2012), la cual se compone de 7 tipos de señales de interés [24], las cuales provienen de 1237 descargas, subdivididas en 201 donde hubo disrupción (señal disruptiva) y 1036 donde no (señal safe). En la Tabla 2-1 se resumen las 7 señales de interés y en la Figura 2-3 se muestra el desarrollo de cada tipo de señal en el tiempo tanto para los casos safe como disruptivo.

Tabla 2-1: Señales de interés JET [24]

Nombre de la Señal	Descripción	Unidad	Nº
Plasma current	Intensidad media de corriente de plasma	A	1
Mode Lock	Densidad de flujo magnético por Amperio	T	2
Plasma internal inductance	Inductancia propia del plasma		3
Plasma density	Densidad electrónica del Plasma	m^{-3}	4
Diamagnetic energy derivative	Derivada de la energía diamagnética almacenada	W	5
Radiated power	Potencia radiada	W	6
Total input power	Potencia total de entrada	W	7

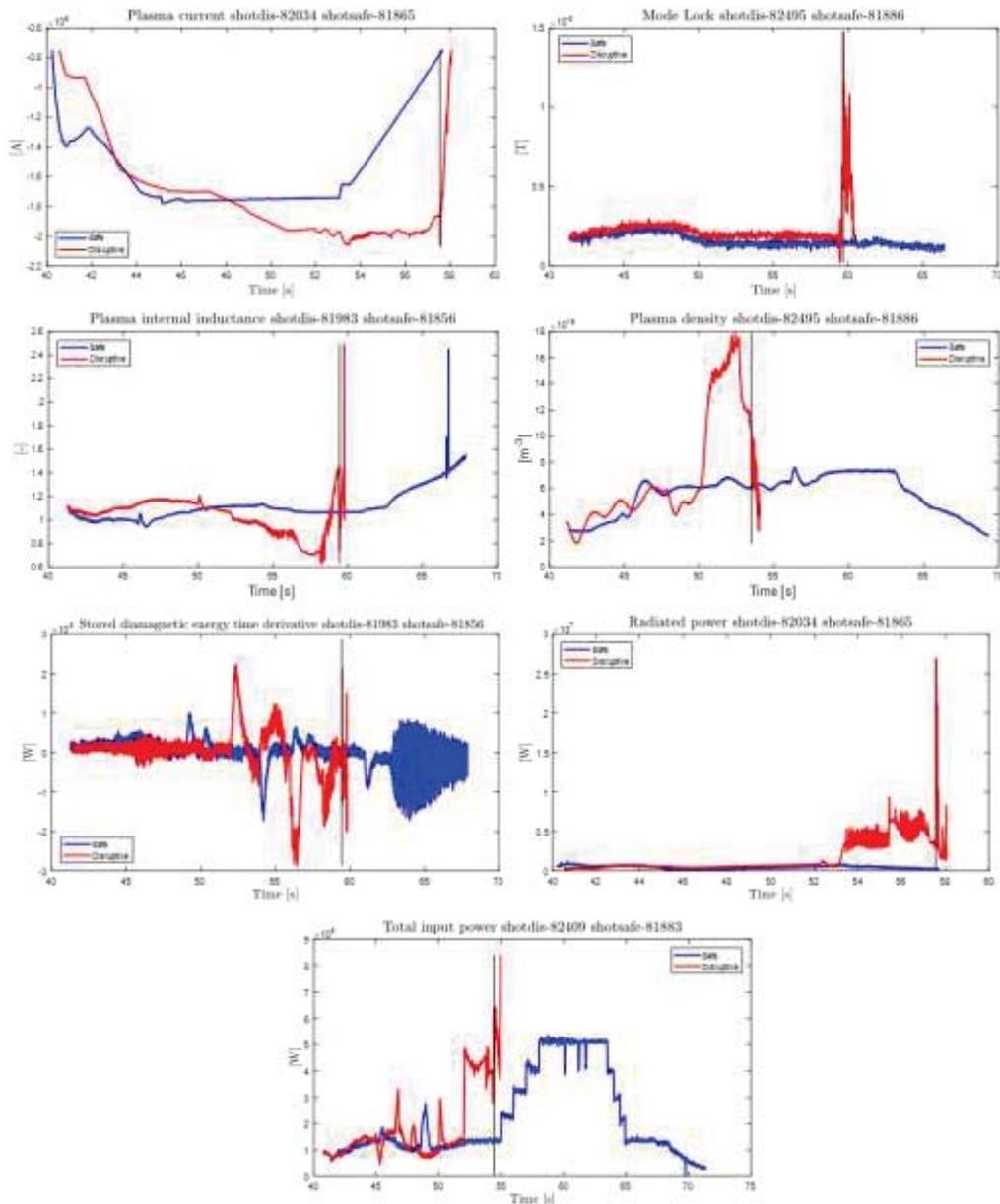


Figura 2-3: Señales de interés Safe y Disruptiva JET

Cabe destacar que cada una de estas señales está muestreada cada 1(ms) y además se eliminaron todas las muestras donde la corriente de plasma no haya alcanzado una amplitud de 750(kA), es decir, que en el caso que en una señal donde la corriente de plasma supera este límite 10(s) después de su inicio, las muestras en este lapso de tiempo son eliminadas en la totalidad de señales, lo que se logra con el procesamiento anterior es eliminar alguna respuesta transitoria del plasma y generar en definitiva información que puede afectar el sistema de predicción, la Figura 2-4 ilustra el modo de recorte de implementado.

Finalmente, a través de estas señales se busca caracterizar el estado disruptivo y no disruptivo del plasma en el dispositivo JET, lo cual típicamente se realiza a partir de técnicas basadas en análisis de datos (aprendizaje automático).

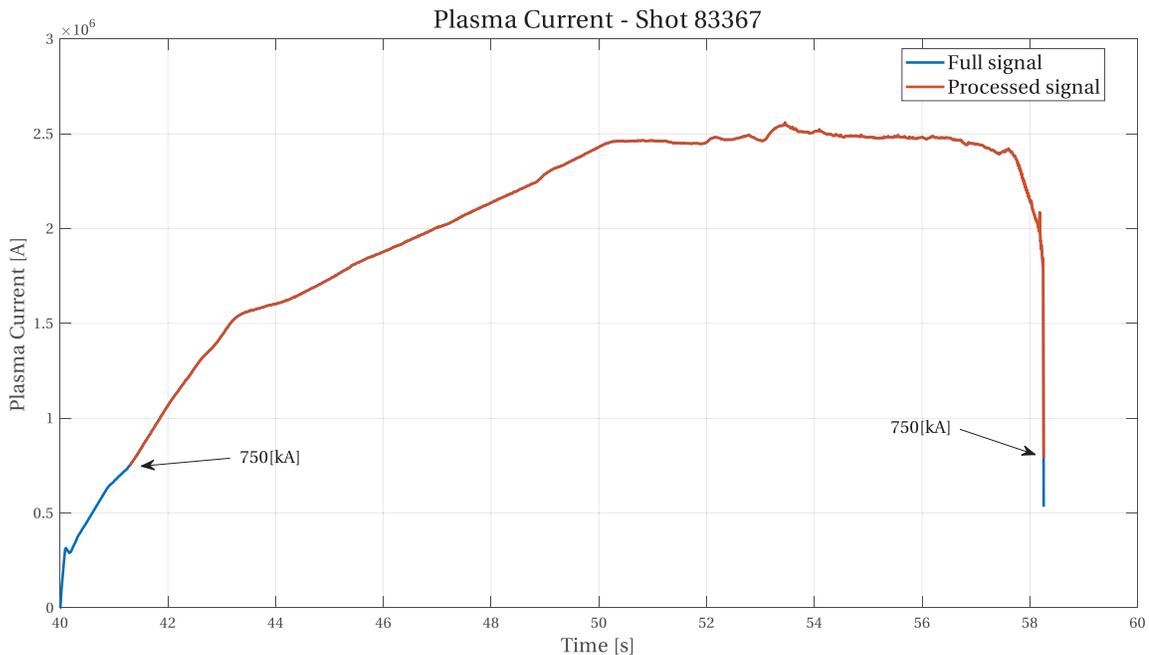


Figura 2-4: Evolución temporal de una descarga JET.

2.3 Enfoque machine learning

Los fenómenos físicos que conducen a estos eventos son muy complejos y no lineales, por lo tanto, se ha abordado esta problemática a través de sistemas de aprendizaje automático, pues éstos no necesitan información respecto a los modelos que definen el comportamiento de las señales, sino que a partir de los mismos datos estos sistemas son capaces de aprender.

2.3.1 Estado del arte

Existe un sinnúmero de investigaciones respecto a la búsqueda de patrones precursores que indiquen el arribo de una interrupción, en donde se destaca principalmente los métodos basados en aprendizaje automático. Cabe destacar que acompañado de la determinación si se producirá el fenómeno disruptivo o no, está la diferencia de tiempos entre el momento donde se genera la interrupción y en cual el fenómeno fue detectado, es decir cuánto tiempo antes el sistema se

anticipa a la interrupción [25], este parámetro de ahora en adelante será nombrado como WarningTime (WT) y se ilustra en la Figura 2-5.

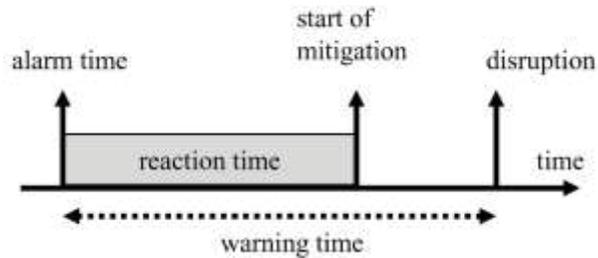


Figura 2-5: Ilustración WarningTime [26]

La siguiente tabla resume alguna de las investigaciones más importantes realizadas con bases de datos del dispositivo JET

Tabla 2-2: Resumen estudio predictores de interrupciones. STD: Desviación Estándar. SVM: Máquina de Vectores Soporte. RNA: Red Neuronal Artificial. SOM: Mapa Auto-organizado. FZ: Lógica Difusa. AD: Detección de Anomalías. PE-SVM: Extensión Probabilística de un SVM. CART: Árboles de Clasificación y Regresión.

Ref.	Autor	Técnica	Acierto (%)	Tiempo Medio antes de Interrupción (ms)	Acierto con tiempo \geq (ms)	STD (ms)	Año
[26]	G. Rattá	SVM	92.7	146,23	-	-	2010
[27]	R. Yoshino	RNA	89.5	-	10	-	2005
[28]	C. Windsor	RNA	86.36	-	10	-	2005
[29]	B. Cannas	RNA/SOM	87.23	-	100	-	2007
[30]	A. Murari	FZ	74	-	100	-	2008
[31]	J. Vegas	AD	89.4	389	-	-	2015
[32]	J.M López	SVM	97	300	-	-	2014
[33]	B. Cannas	SVM	91	-	-	-	2007
[34]	G. Pautasso	RNA	79	-	-	-	2002
[35]	A. Maruri	PE-SVM	97.3	316	-	333	2018
[36]	J. Vegas	VP	94	654	-	-	2014
[37]	M. Lungarona	CART	96.48	271	-	342	2018

Es importante destacar que el objetivo que ha planteado ITER para el desarrollo de modelos se establece como una tasa de verdaderos positivos (todos aquellos clasificados como disruptivos y que realmente lo son) mayor o igual a 95(%), falsos positivos (catalogados como no-disruptivo siendo disruptivos) de no más de 5% y los tiempos de predicción con un mínimo de anticipación de 30(ms), aunque esto último ha bajado a 10(ms) por los avances en las técnicas de mitigación. De acuerdo a lo anterior planteado y según los artículos estudiados en la Tabla 2-2, se refleja la necesidad mejorar el sistema de predicción con finalidad de obtener mejores tasas de acierto y tiempos de anticipación a la interrupción, adicionalmente otro aspecto importante tiene relación con la complejidad del modelo, es decir que el sistema tenga un enfoque del tipo caja transparente, siendo esto último una de las principales motivaciones del sistema propuesto del presente trabajo.

2.3.2 Sistema de caja negra

El enfoque caja negra es utilizado para representar sistema que son estudiados tan solo de las perspectivas de la entrada y la salida, es decir, solo es importante que salida obtengo para una entrada determinada y no se toma en cuenta el funcionamiento interno de éste. La Figura 2-6 ilustra un esquema básico de un sistema de caja negra.



Figura 2-6: Ilustración básica Sistema con enfoque caja negra

Cuando los modelos tienen comportamientos altamente complejos y con una gran cantidad de variables, es donde generalmente se implementa este enfoque. Por otro lado, cuando este tipo de orientación se aplica en sistema de clasificación, lo que se busca es forzar el comportamiento del sistema para obtener una salida deseada. Un ejemplo clásico son las redes neuronales artificiales, las cuales pueden tener una arquitectura conformada por cientos de miles de parámetros, que a su vez no tienen una interpretación clara.

2.3.3 Sistema de caja transparente

Por el contrario al enfoque anterior, existen los sistemas de caja transparente, en los cuales si es posible tener un conocimiento respecto al funcionamiento interno del proceso. En este caso la cantidad de variables es más reducida y además cada una de ellas es interpretable lógicamente, es decir que pueden ser alteradas y saber de antemano como afecta al sistema y en sus salidas. La Figura 2-7 muestra un ejemplo de un sistema bajo este enfoque a través de un circuito eléctrico básico.

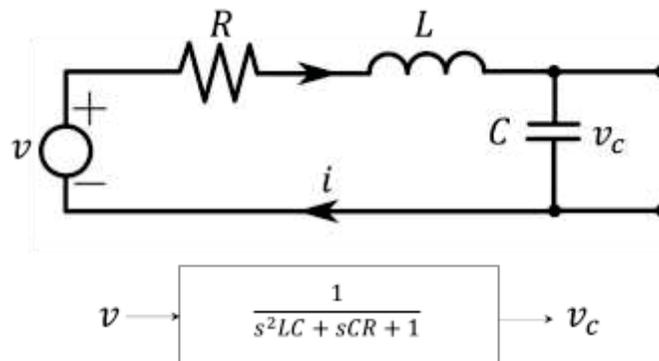


Figura 2-7: Ejemplo circuito RLC serie, sistema caja transparente

Este concepto aplicado en un sistema predictor de disrupciones puede ser una herramienta muy importante, ya que de esta manera es posible modificar parámetros del sistema y ajustarlos a las necesidades particulares de una descarga, por ejemplo, aumentar su sensibilidad ya que la descarga será de alta potencia y una disrupción provocaría grandes daños.

2.4 Objetivos del proyecto

De acuerdo a la problemática planteada se proponen los siguientes objetivos clasificados en generales y específicos.

2.4.1 Objetivo general

Diseñar e implementar un sistema de predicción de disrupción mediante detección de anomalías.

2.4.2 Objetivos específicos

1. Estudiar la problemática de la disrupción en dispositivos de fusión nuclear.
2. Estudiar enfoques de la detección de anomalías y su implementación en predicción de disrupciones.
3. Diseñar algoritmos de detección de disrupción mediante detección de anomalías.
4. Implementar los algoritmos diseñados y probarlos con bases de datos reales de un dispositivo de fusión nuclear.

3 Detección de anomalías

3.1 Introducción

Detección de anomalías se refiere a la búsqueda de patrones en los datos que no se ajustan al comportamiento común o esperado de dichos datos, estos patrones son denominado habitualmente como anomalías (anomalía), outlier (valores atípicos), exceptions (excepciones), aberrations (aberraciones), surprises (sorpresas), peculiarities (peculiaridades), siendo las dos primeras las más comunes. En [38], se define un outlier como “una observación que se desvía mucho de otras observaciones y despierta sospechas de ser generada por un mecanismo diferente”. Este enfoque posee una amplia gama de aplicaciones, en las cuales se encuentran, fraudes en tarjetas de crédito, seguridad cibernética, vigilancia militar, medicina, daños industriales y redes de sensores.

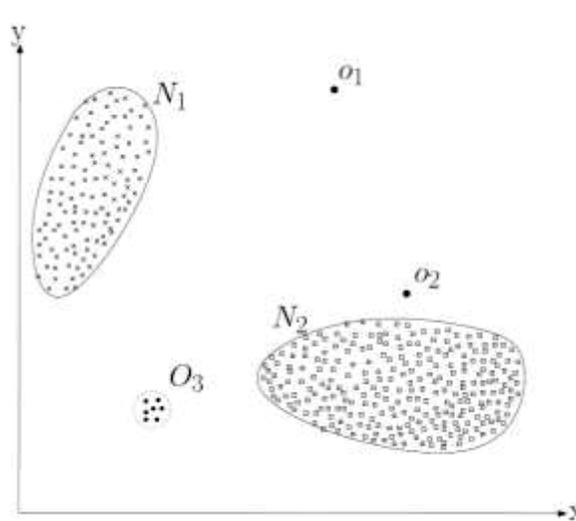


Figura 3-1: Ejemplo ilustrativo para la detección de anomalías en dos dimensiones [36]

Un ejemplo ilustrativo se muestra en la Figura 1-2 a través de un conjunto de datos de dos dimensiones, en los cuales se observan dos regiones “normales”, N_1 y N_2 , ya que la mayoría de los datos se encuentran en dichas regiones. Por otro lado los puntos o_1 y o_2 , y la región O_3 pueden ser considerados como outlier.

Existe una gran cantidad de métodos para la detección de outlier, los cuales se pueden aplicar de acuerdo a la naturaleza que poseen los datos. Una muestra se puede describir como algún

conjunto de atributos o características, las cuales a su vez pueden ser binarias, continuas o categorizadas. La cantidad de atributos de la muestra determina si se está en el caso univariantes o multivariantes, por ejemplo, la muestra o_2 claramente se trata de una anomalía, pero si se analiza respecto a un único eje “x” o “y” (univariantes), no se consideraría de dicha manera.

3.2 Tipos de anomalías

El tipo de anomalía buscado define también el método a emplear, típicamente se clasifican en tres categorías:

3.2.1 Anomalías puntuales

Es el enfoque más utilizado en el área de la detección de anomalías. Una muestra se considera anómala cuando se encuentra apartada de lo considerado como normal, los puntos o_1 y o_2 de la Figura 3-1 son consideradas como anomalías puntuales.

3.2.2 Anomalías colectivas

Cuando un conjunto de muestras es considerado anómalo y no una muestra única por sí misma, estamos en presencia de anomalías colectivas, es decir que una muestra individual en un conjunto anómalo puede no ser considerada como outlier. Por ejemplo, de acuerdo al electrocardiograma mostrado en la Figura 3-2, la sección resaltada de la señal indica una anomalía, debido a la duración en que la señal está con valores bajos, pero dichos valores bajos por sí mismos no son valores atípicos.

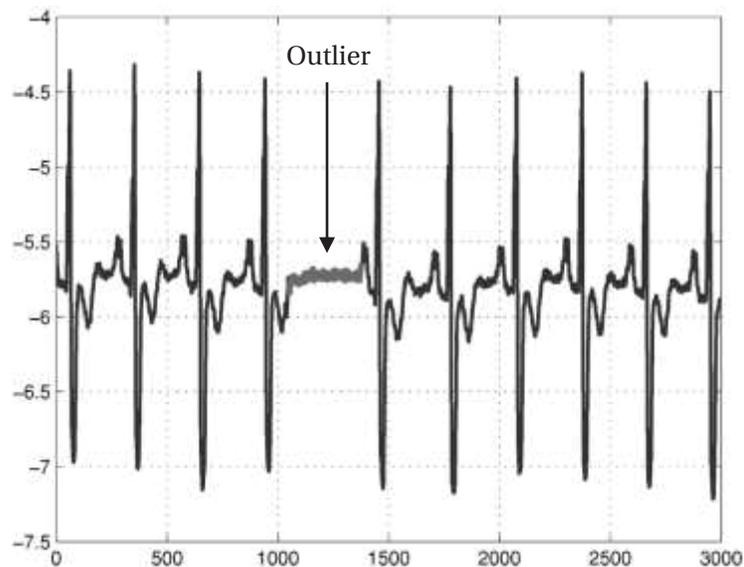


Figura 3-2: Ejemplo ilustrativo para anomalías colectivas en un electrocardiograma [36]

3.2.3 Anomalías contextuales

Si una muestra es anómala bajo un contexto específico, pero no de otra manera, es considerada como anomalía contextual o condicional. El primer acercamiento al contexto de las muestras

viene dado por su estructura y debe ser abordado en la formulación del problema. Cada observación de los datos puede ser definida utilizando dos conjuntos de características: Atributos Contextuales, determinan el contexto para cada instancia, por ejemplo, pueden ser espaciales o temporales; Atributos de Comportamiento, muestran la manera que se comporta la muestra, por ejemplo, temperatura (comportamiento) en una región específica (contexto). La Figura 3-3 muestra un ejemplo ilustrativo que grafica este tipo de anomalías, en donde se observa que la temperatura t_2 tiene un comportamiento fuera de lo común debido al mes en que se genera, es decir, su contexto. Cabe señalar que la temperatura t_2 por sí mismo no es anómala, ya que este mismo valor fue medido en t_1 donde es considerada como una muestra normal.

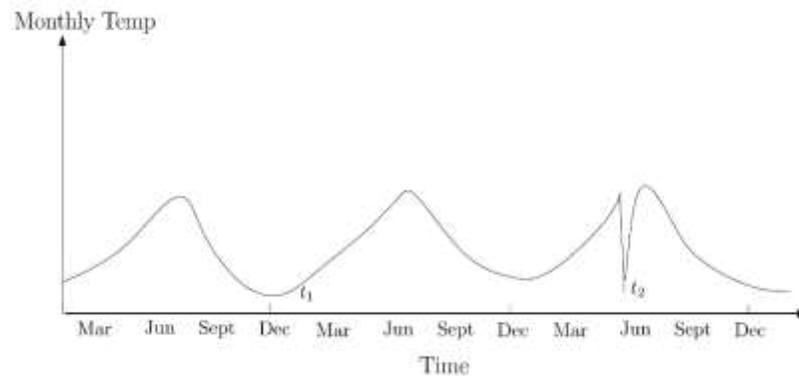


Figura 3-3: Ejemplo ilustrativo para anomalías contextuales en temperatura mensual [36]

3.3 Etiqueta de los datos

Las etiquetas en los datos están relacionadas con el comportamiento normal o anómalo de las muestras, en este punto se pueden destacar tres modos: Supervisado, Semisupervisado y No Supervisado, en el primero se tiene la disponibilidad de la etiqueta tanto de las observaciones normales como las anormales, típicamente esto se realiza a partir del criterio de un experto; para el segundo modo se tiene solo el etiquetado de las muestras que se consideran como normales y en el enfoque No Supervisado las etiquetas no existen, los métodos ocupados en este caso suponen que las instancias normales son bastante más frecuentes que las anómalas.

Es importante destacar dos fundamentales problemas respecto al etiquetado de las muestras, típicamente conseguir un conjunto de datos etiquetados lo suficientemente robustos para cubrir cualquier tipo de anomalías es muy complejo y sumado a lo anterior se tiene que el comportamiento dinámico de las anomalías puede provocar su no detección.

3.4 Métodos de detección

No existe un método de detección de anomalías lo suficientemente general para abarcar cualquier problema, por esta razón la mayoría de las técnicas empleadas son formuladas para un tipo de problema específico. En la literatura se pueden encontrar métodos generados para problemas muy específicos y otros más generales, en la Figura 3-4 se clasifican y resumen algunos métodos existentes.

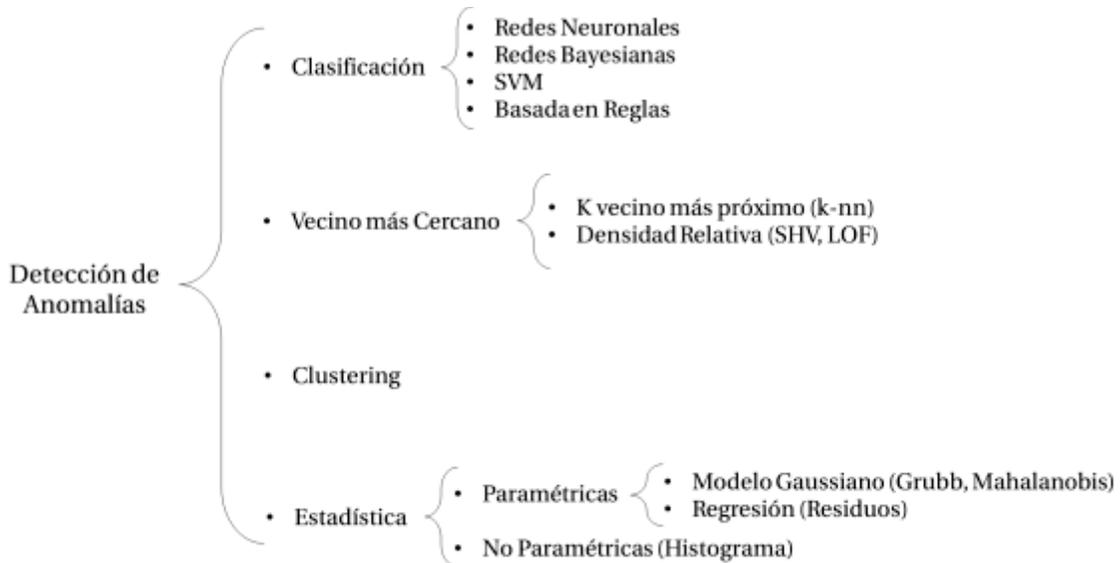


Figura 3-4: Clasificación de métodos para la detección de anomalías

3.4.1 Clasificación

A través de datos etiquetados se entrena un clasificador, el cual puede ser para una clase única o multiclase. En este caso se supone que existen clasificadores capaces de distinguir clases normales y anómalas de acuerdo a las características dadas.

3.4.2 Vecino más cercano

La suposición para este caso es que los datos normales se encuentran en vecindades más densas en comparación con muestras anómalas. Este tipo de técnicas requieren una medida de distancia, usualmente se utiliza la Euclidiana, pero se pueden utilizar otras.

3.4.3 Clustering

Este tipo de técnicas busca el agrupamiento de datos que son similares, generalmente se realiza de forma no supervisada, pero existen técnicas semisupervisadas. Se puede dividir en tres subcategorías que se basan en las siguientes suposiciones:

1. Categoría 1: Las muestras normales pertenecen a una agrupación de datos, mientras las anómalas no son parte de ningún grupo.
2. Categoría 2: Las muestras normales se encuentran cerca del centro de gravedad del grupo más cercano, mientras que los outlier se encuentran lejos.
3. Categoría 3: Los datos normales pertenecen a grupos densos y grandes, mientras que los anómalos en grupos dispersos y pequeños.

Al igual que las técnicas del Vecino más Cercano son necesarias las medidas de distancia, sin embargo, la diferencia radica en que en clustering se evalúa cada caso respecto al grupo que es parte, en cambio las técnicas basadas en el vecino más próximo las muestras se analizan respecto a su entorno local.

3.4.4 Estadística

Estos tipos de métodos se basan bajo la premisa que los datos catalogados como normales se producen en zonas de alta probabilidad en un modelo estocástico y por otro lado las instancias anómalas se encuentran en regiones de baja probabilidad. Por un lado, existen las técnicas paramétricas, que suponen un modelo de densidad de probabilidad y parámetros que la definen para las muestras normales y las no paramétricas, donde el modelo no se define a priori.

Es importante destacar que no existe una clasificación establecida para los métodos de detección de outlier, el que se presentó anteriormente es basado en [39], pero en [40] se muestra otra clasificación referido a la detección de Fraude.

3.5 Detección de anomalías para la predicción de interrupciones

Un estudio respecto a trabajos que relacionan la detección de anomalías y interrupciones puede ser favorable para posibles futuros estudios, a partir de esto se seleccionó un artículo denominado “*Real-time anomaly detection for disruption prediction: the JET case*”, donde se estudió la implementación de un predictor en base a la detección de Outlier a partir de bases de datos obtenidas del dispositivo JET [31].

Como primera parte se acordó un análisis de acuerdo a ventanas de muestras con un ancho de 32(ms) (las muestras originales se encuentran cada 1(ms)), esta ventana a su vez se desplaza cada 2(ms) como se ilustra en la Figura 3-5.

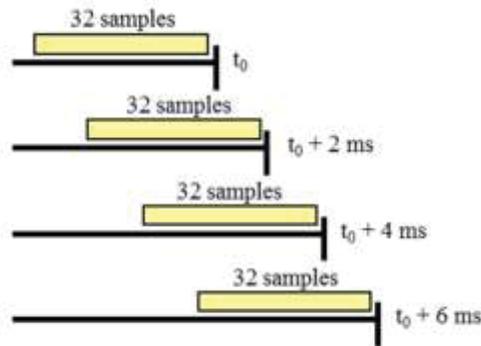


Figura 3-5: Esquema de ventanas deslizantes [31]

A continuación, a cada ventana de 32(ms) se le realiza un mapeo a través de la familia Haar Wavelet [41] ($f: \mathbb{R}^{32} \rightarrow \mathbb{R}^p$) de acuerdo a diferentes niveles de descomposición. Así por ejemplo para una descomposición de nivel 4 el mapeo sería $f: \mathbb{R}^{32} \rightarrow \mathbb{R}^2$, disminuyendo la dimensionalidad del problema. La elección de este mapeo con la transformada wavelet se discute en [31]. Una vez obtenida la muestra es posible graficarla en un entorno de dos dimensiones como se muestra la Figura 3-6.

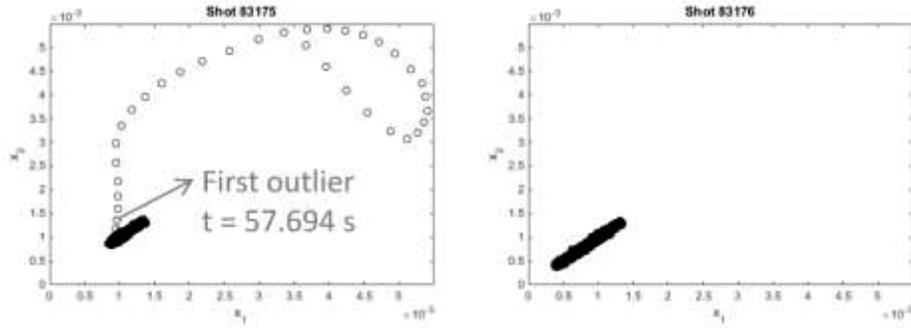


Figura 3-6: Espacio de características cada 2(ms) en señal disruptiva y no-disruptiva [31]

Ahora la tarea es determinar cuándo las muestras que se escapan de lo común y predecir la disrupción. Para esto la distancia euclidiana no hace un buen trabajo debido a que sus perímetros de igual distancia son círculos y por lo tanto no toma en cuenta el patrón de covarianza de los datos, por lo mismo la métrica utilizada se realizó a través de la distancia Mahalanobis [42]. Finalmente se establece el criterio que se muestra en la ecuación (3-1).

$$\left| \frac{DM(t_p) - \text{media}(DM(t < t_p))}{\text{std}(DM(t < t_p))} \right| \geq K_M \quad (3-1)$$

Donde K_M corresponde al umbral de detección que fue determinado empíricamente con un valor de 10, en definitiva si este umbral se sobrepasa se gatilla la alarma disruptiva, en la Figura 3-7 se ilustra lo anterior de manera gráfica.

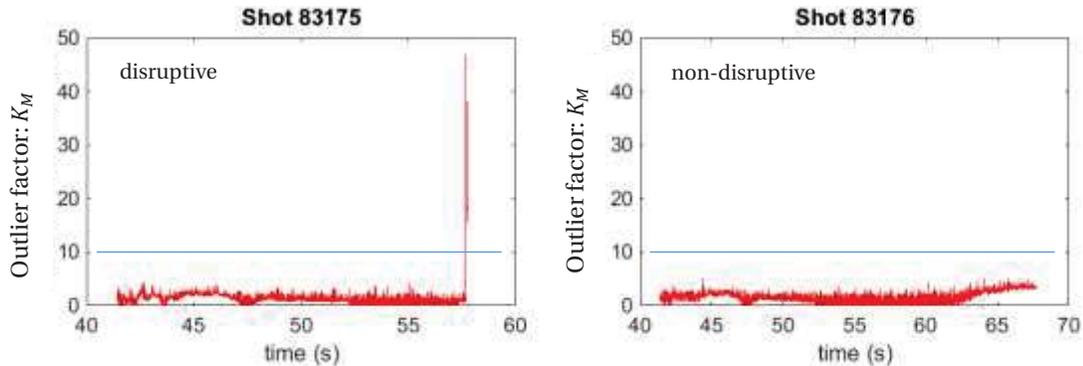


Figura 3-7: Evolución temporal del criterio de alarma de dos diferentes descargas [31]

Esta explicación se realizó a partir de mapeo a dos dimensiones, pero se puede extrapolar para los niveles de descomposición de la transformada wavelet 1, 2 y 3 ($f: \mathbb{R}^{32} \rightarrow \mathbb{R}^{16}$, $f: \mathbb{R}^{32} \rightarrow \mathbb{R}^8$ y $f: \mathbb{R}^{32} \rightarrow \mathbb{R}^4$, respectivamente). Finalmente, los resultados obtenidos alcanzaron una tasa de verdaderos positivos de 89.4(%) y falsos positivos del 8.98(%), lo cual está relativamente distante a los requerimientos establecidos por ITER.

Se concluye a través de este estudio que la detección de anomalías puede generar grandes avances respecto a la predicción de disrupciones, debido a su sencillez de funcionamiento, el amplio margen de posible mejora en los resultados y que para este enfoque existe una amplia gama de alternativas que pudiesen ser aplicadas, lo cual converge finalmente en una gran motivación para su implementación.

4 Predictor de Disrupciones basado en Detección de Anomalías (DPAD)

4.1 Lag-plot - Mapa autorregresivo

Lag-plot (gráfica de retraso en español) o mapa autorregresivo, hace referencia a una herramienta matemática utilizada para la representación de datos o series temporales, ayudando a evaluar la existencia de aleatoriedad o de algún patrón, también este tipo de representación es aplicado para ayudar a identificar valores anómalos (outliers).

4.1.1 Definición

Para valores de datos Y_1, Y_2, \dots, Y_N , el retardo de periodo k del valor Y_i se define como el valor del dato ocurrido k muestras antes de la muestra i . Es decir:

$$Lag_k(Y_i) = Y_{i-k} \quad (4-1)$$

Por ejemplo para $Lag_1(Y_2) = Y_1$ y $Lag_4(Y_{12}) = Y_8$.

Para realizar la gráfica a través de la función $Lag_k(Y_i)$ (Lag-plot), en un plano cartesiano se grafican los pares ordenados de la forma $(Lag_k(Y_i), Y_i)$. En particular el análisis planteado es a través de muestras consecutivas, lo cual implica que k toma el valor de 1.

4.1.2 Patrones lagplot

La herramienta Lag-plot es capaz de ayudar en la identificación de patrones posiblemente ocultos en los datos o series temporales, en esta sección se muestran las tres principales estructuras de los datos a través de esta herramienta.

Datos aleatorios

Este tipo de distribución en los datos da a lugar a una representación sin patrón al utilizar la herramienta Lag-plot, los datos están distribuidos de manera dispersa tanto para el eje de las ordenas como de las abscisas, la Figura 4-1 ilustra lo anterior gráficamente.

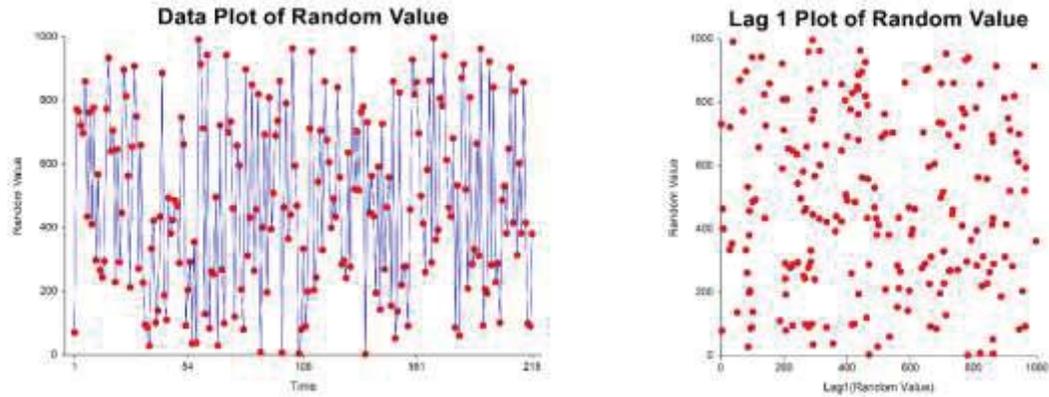


Figura 4-1: Lag-plot datos aleatorios (fuente: <https://www.ncss.com>)

Datos con autocorrelación

Los datos con Autocorrelación generan gráficas con patrones lineales que siguen la diagonal (recta de pendiente 1 que corta el eje “x” e “y” en el valor 0). Mientras más alto es el nivel de Autocorrelación de los datos, los puntos en la gráfica se agrupan de manera más densa a lo largo de la diagonal. Las siguientes figuras muestran los mapas de auto regresión para datos con menor y mayor autocorrelación.

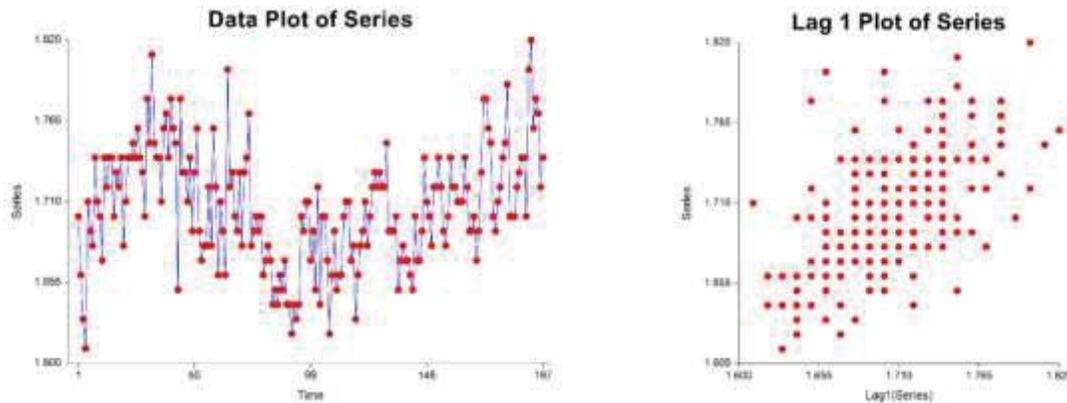


Figura 4-2: Lag-plot datos con autocorrelación débil (fuente: <https://www.ncss.com>)

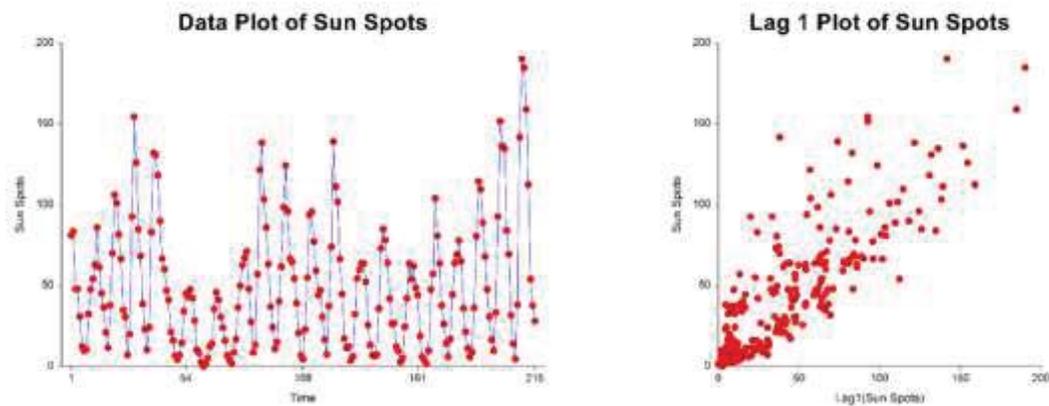


Figura 4-3: Lag-plot datos con autocorrelación moderada (fuente: <https://www.ncss.com>)



Figura 4-4: Lag-plot datos con autocorrelación fuerte (fuente: <https://www.ncss.com>)

Datos sinusoidales

Los datos sinusoidales forman gráficas de retraso con patrones de estructura circulares o elípticas como se muestra en la Figura 4-5.

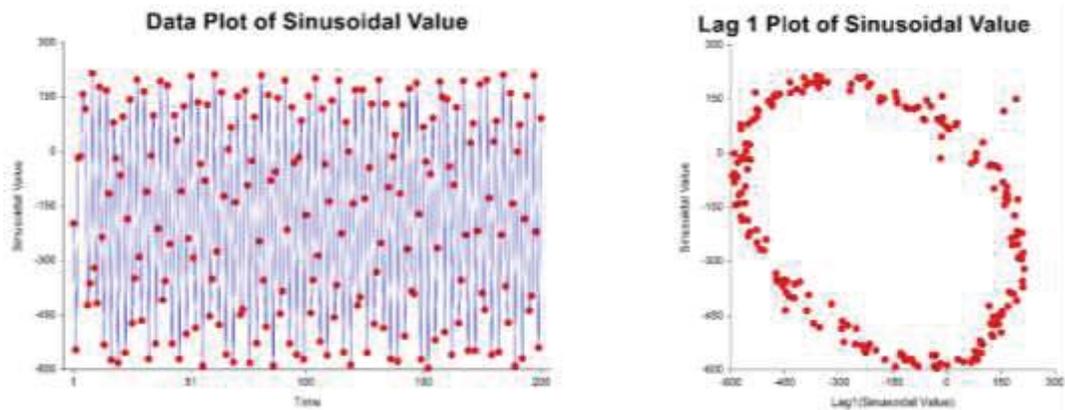


Figura 4-5: Lag-plot datos Sinusoidales (fuente: <https://www.ncss.com>)

En la figura anterior se observan claramente dos muestras apartadas de la elipse conformada por los datos, dichas muestras pueden ser consideradas como anómalas. Además, es importante destacar que en la gráfica temporal estas anomalías no se observan de manera tan evidente. Esta última forma de presentar los datos es la empleada para la construcción del predictor de disrupciones mediante detección de anomalías.

4.1.3 Lag-plot base de datos

En las siguientes figuras se muestran los Lag-plot o mapas autorregresivos de los diferentes tipos de señales que conforman la base de datos del dispositivo JET, tanto para los disparos safe como los disruptivos.

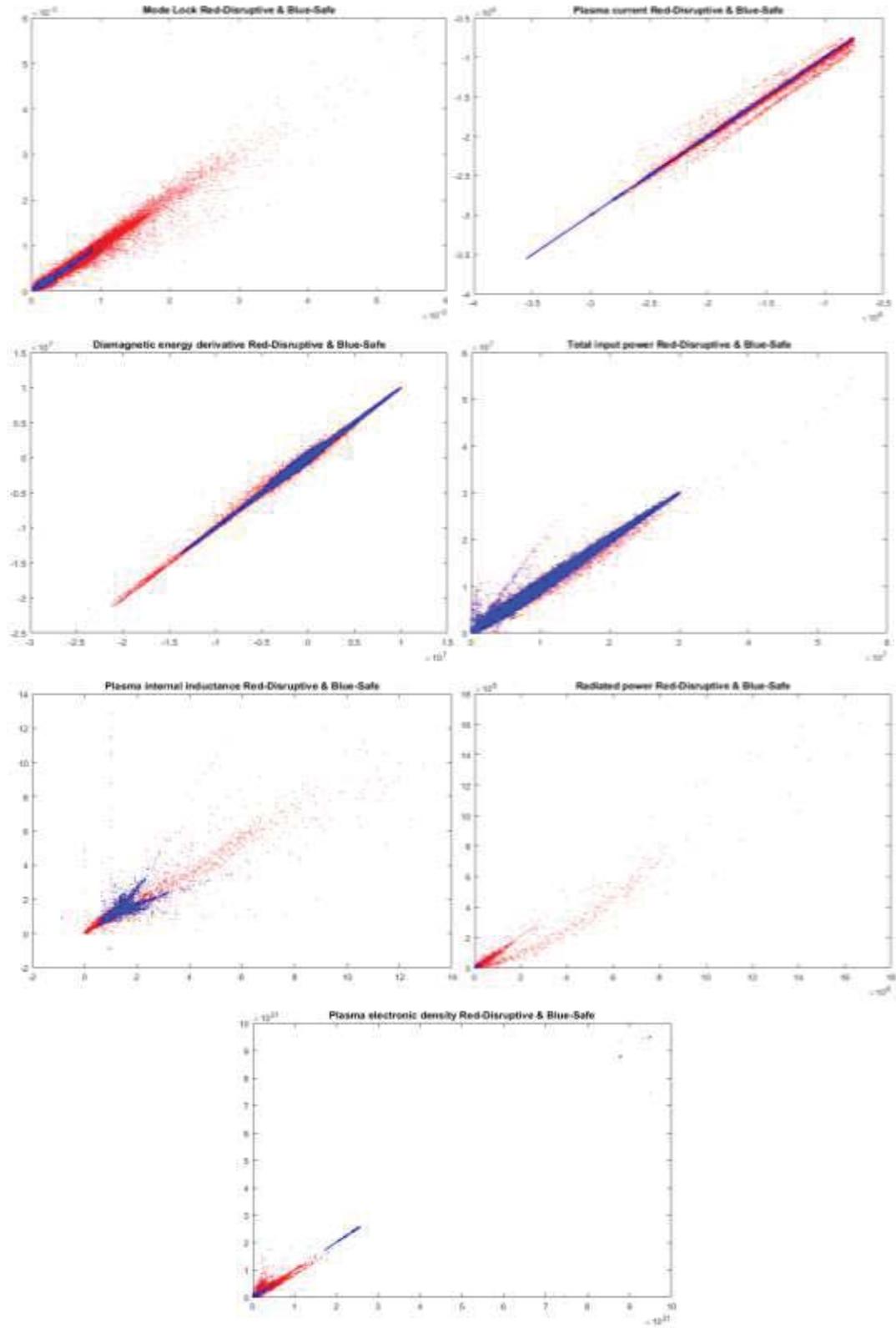


Figura 4-6: Lag-plot de señales de interés de toda la base datos de disparos safe y disruptivos (eje "x" corresponde a la muestra $t - 1$ y el eje "y" a la muestra t)

A partir del análisis de las figuras anteriores, es posible observar que existen una mayor dispersión de las muestras en los datos disruptivos que en las descargas safe para algunos tipos de señal, por ejemplo, es más factible detectar outlier precursores de disrupciones para señales Mode Lock en contraste a las señales de potencia total de entrada (Total input power).

4.2 Detección de anomalías

En particular respecto a las señales a utilizar, las anomalías son del tipo puntal, en segundo término el etiquetado de las señales (safe o disruptivo) se puede relacionar como información privilegiada, aunque mirado desde el punto de vista interno de cada señal, no se sabe si una muestra por sí misma (amplitud y tiempo específicos) es anómala o no (es precursora de disrupción o no) y adicionalmente, la suposición que se relaciona de mejor manera con la naturaleza de los datos tiene que ver con la detección de outlier a través de un clasificador basado en regla (superior a un umbral se detecta un outlier).

En la Figura 4-7 se muestra un ejemplo ilustrativo mediante las señales Mode Lock para todos los disparos disruptivos y safe, donde se establece un umbral que divide outliers con muestras normales.

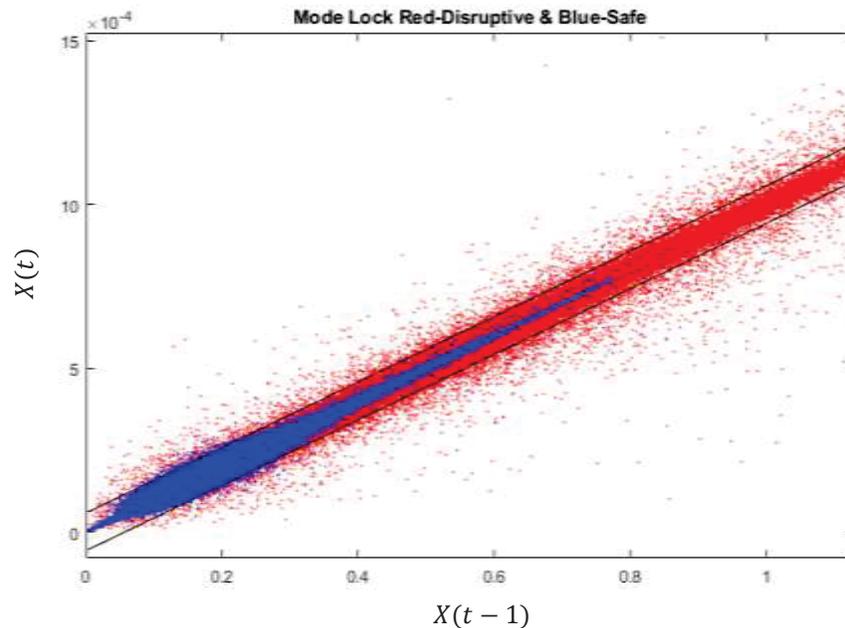


Figura 4-7: Ejemplo Lag-plot Mode Lock y umbral detector de anomalías

4.3 Arquitectura predictor

En esta sección se desarrollará la explicación del predictor de disrupciones propuesto en sus dos etapas, describiendo cada una de las fases con que se componen junto a los distintos parámetros establecidos para su funcionamiento y como cada uno de ellos afecta el desarrollo del sistema predictor.

4.3.1 Etapa 1

Detección de outlier

El sistema de predicción de disrupciones planteado se basa en la detección de outlier dentro de cada tipo de señal, donde las muestras son catalogadas como anómalas cuando la diferencia de la amplitud de una muestra específica con la inmediatamente anterior supera un umbral establecido, lo anterior se puede expresar matemáticamente como se muestra en (4-2).

$$outlier_{it} = \begin{cases} 1, & |X_{ij}(t) - lag_1(X_{ij}(t))| > \sigma_i * d_i \\ 0, & eoc \end{cases} \quad (4-2)$$

Donde $outlier_{it}$ es un vector que toma valores de 1 si se genera un outlier o 0 en otro caso, para el tipo de señal i (correspondiente N° de tipo de señal indicado en la Tabla 2-1) en el instante t , $X_{ij}(t)$ es la serie temporal de la señal tipo i del disparo j en el instante t , σ_i corresponde a la desviación estándar de las diferencias de muestras consecutivas de todos las descargas safe con que se compone la base de datos, para el tipo de señal i y se puede expresar matemáticamente según la ecuación (4-3).

$$\sigma_i = DesviaciónEstandar \left(|X_{ij}(t) - lag_1(X_{ij}(t))| \right), \forall t \wedge \forall j \in \{Safe\} \quad (4-3)$$

Por otro lado d_i corresponde a un parámetro que ajusta la detección de outlier o ajuste de umbral para cada tipo de señal i , de manera de poder disminuir o aumentar el umbral (anchura del tubo de detección), éste puede tomar valores dentro del rango $[0,3]$. Cabe destacar que dicho límite fue establecido bajo el concepto de que en un rango $\pm 3 * \sigma_i$ respecto a la media se abarcan aproximadamente el 99.73(%) de las muestras. La detección de una muestra anómala de acuerdo al umbral se ilustra en la Figura 4-8.

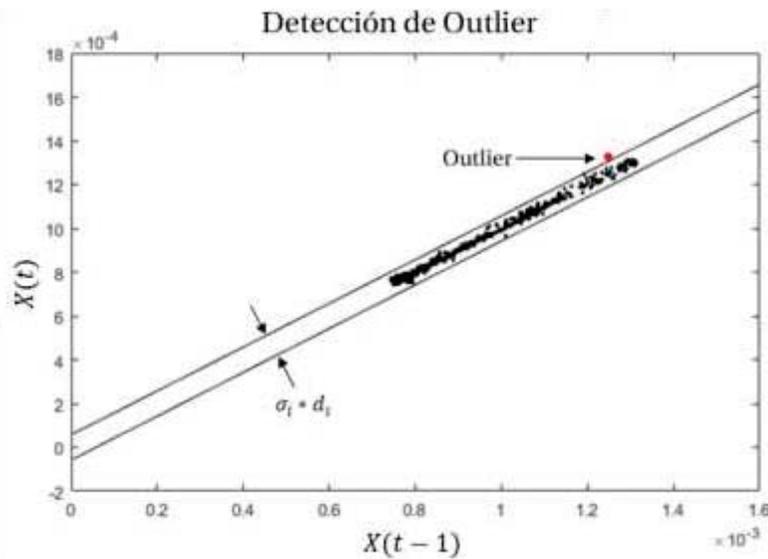


Figura 4-8: Ilustración de detección de anomalía propuesto a través de umbral

El sistema de predicción evalúa la aparición de anomalías en cada tipo de señal, de manera que al no detectarse una muestra como outlier el predictor continua con la siguiente muestra, de manera contraria avanza con los siguientes criterios de alarma disruptiva

Criterio 1 – Anomalías simultáneas

En el caso de existir una anomalía en una de las siete señales en el instante t se continua con el primer criterio de alarma, el cual se expresa matemáticamente de acuerdo a (4-4).

$$Criterio_{1t} = \begin{cases} True, & \sum_{\forall i} outlier_{it} \geq \alpha \\ False, & e. o. c \end{cases} \quad (4-4)$$

El criterio se vuelve verdadero cuando se hayan detectado un número igual o mayor de α outlier en los 7 tipos de señales en un instante t determinado y en dicho caso se lanza la alarma de disrupción. El parámetro α puede tomar un valor del conjunto $\{1, 2, 3, 4, 5, 6, 7\}$ y se puede definir como la cantidad de outlier necesarios en un mismo instante y en los distintos tipos de señales para generar la alarma de disrupción, de manera que cuando α toma valores bajos, el criterio de alarma se vuelve menos exigente, aumentando la tasa de Falsos Positivos (disparos no disruptivos donde se haya lanzado la alarma) y en caso contrario, para valores altos el sistema se vuelve más estricto generando un perjuicio a la tasa de Verdaderos Positivos (disparos disruptivos donde se haya lanzado la alarma), en definitiva la asignación de α es sensible al comportamiento del sistema predictivo. En la Figura 4-9 se ilustra mediante un ejemplo el primer criterio de alarma descrito.

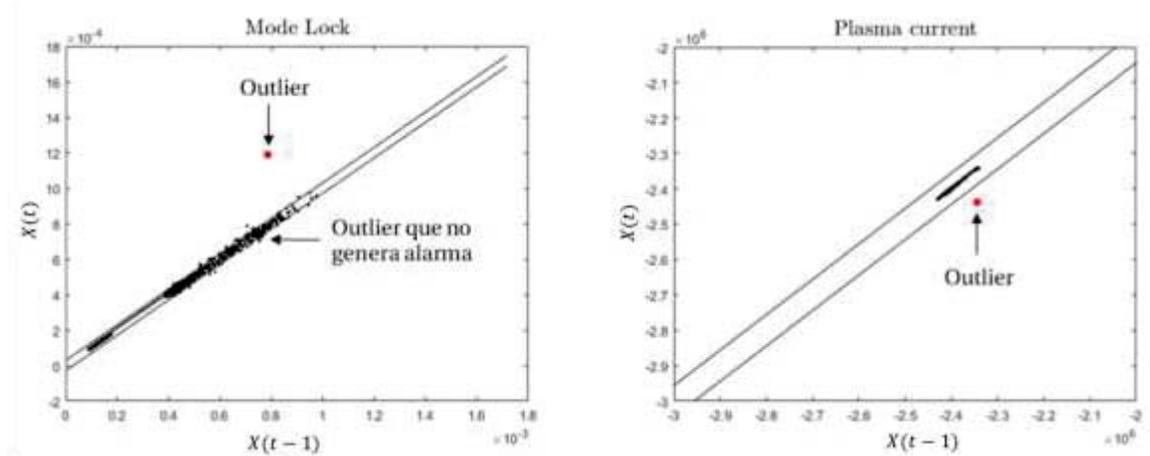


Figura 4-9: Ejemplo gráfico primer criterio de alarma

Es importante destacar que se generan outlier en la señal Mode Lock de la Figura 4-9 pero que no gatillan alarma de disrupción debido a que en ese mismo instante en otro tipo de señal no se produjo una muestra anómala, los outlier destacados en rojo si suceden en el mismo instante por lo que para un valor de α igual a 2 se dispararía la alarma disruptiva.

Criterio 2 – Anomalías consecutivas

Si el proceso anterior no lanza la alarma el sistema evalúa el siguiente criterio de disrupción, en este caso se analizan las muestras anómalas que suceden de forma consecutiva en un mismo tipo de señal, esto se puede formular de acuerdo a la expresión (4-5).

$$Criterio_{2t} = \begin{cases} True, & \sum_{k=t+1-\beta}^t outlier_{ik} \geq \beta, \forall i \\ False, & e. o. c \end{cases} \quad (4-5)$$

Este último criterio se vuelve verdadero y se dispara la alarma de disrupción cuando β muestras anómalas suceden de forma consecutiva en un mismo tipo de señal, en este caso el parámetro β puede tomar cualquier valor entero positivo ($\beta \in \mathbb{N}$) y a medida que su valor aumenta la exigencia del criterio también lo hace (de la misma forma que el primer criterio). En la Figura 4-10 se grafica el funcionamiento de este criterio.

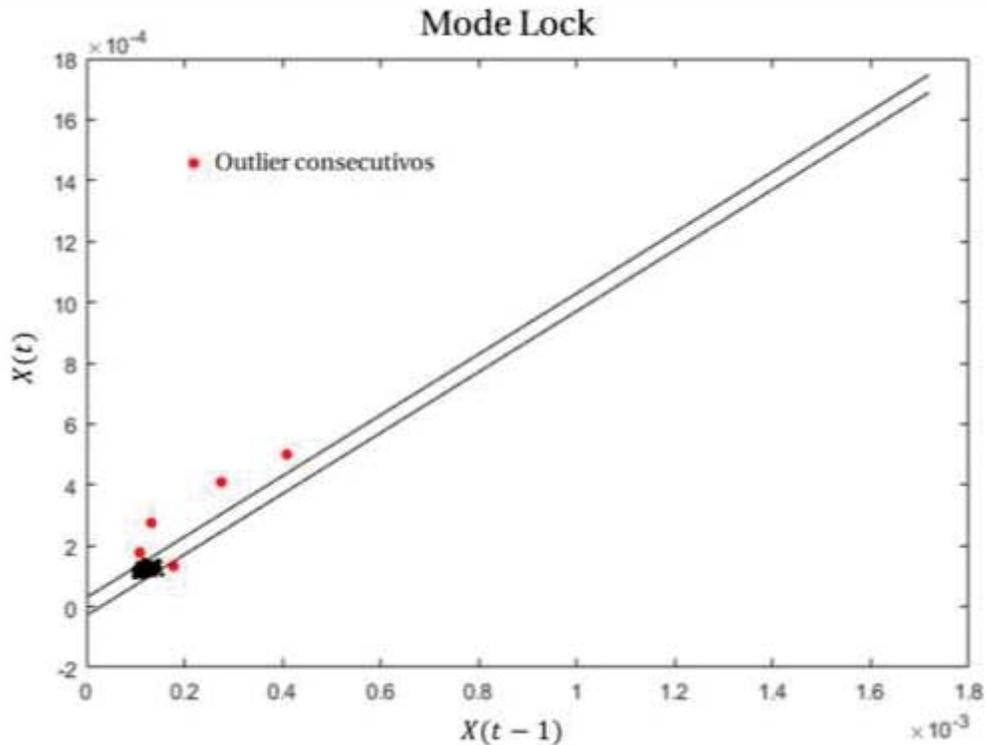


Figura 4-10: Ejemplo gráfico segundo criterio de alarma

En la gráfica de la Figura 4-10 se muestra como muestras anómalas que no suceden de forma consecutiva no provocan el disparo de la alarma, pero cuando esto sucede β veces seguidas si se dispara la alerta disruptiva, en este ejemplo ilustrativo β tendría un valor igual a 5.

4.3.2 Diagrama de bloques etapa 1

En esta sección se muestra el sistema de predicción descrito en el apartado anterior a través de un diagrama de bloques.

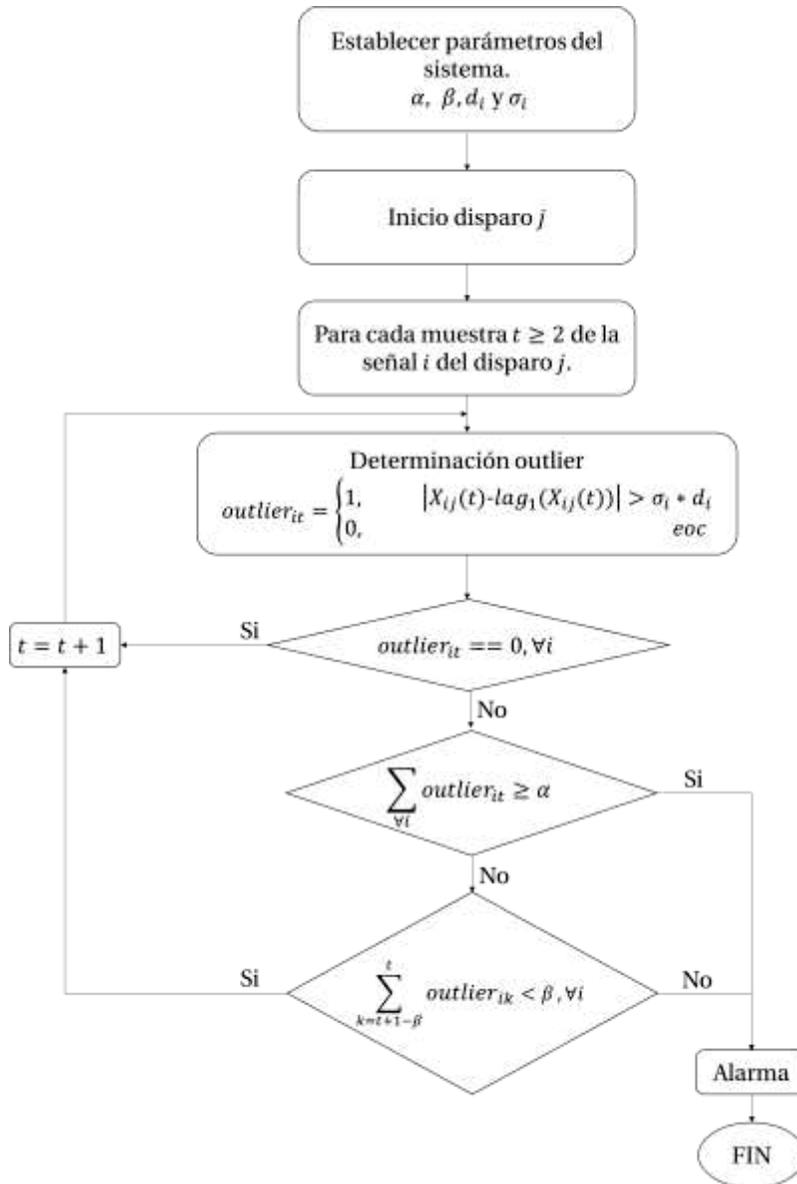


Figura 4-11: Diagrama en bloques del sistema de predicción propuesto en su primera etapa

El modo de entrenamiento del sistema DPAD en su primera etapa será descrito en el siguiente capítulo.

4.3.3 Etapa 2

Para mejorar la efectividad del sistema DPAD se realizaron una serie de transformaciones respecto a su primera etapa, en donde principalmente destaca la inclusión de un nuevo criterio de alarmas y la modificación del parámetro d_i .

Modificación de parámetros

Lo que se plantea en este punto, es realizar una modificación de los parámetros del modelo, para ello se optó por la separación del ajuste de umbral d_i , de manera que cada umbral se acomode de mejor manera a un criterio en particular y así obtener una mejor resultado del sistema predictor. Entonces, d_i cambiaría su subíndice a d_{ih} , donde i corresponde al tipo de señal y h al criterio de evaluación asociado y por consiguiente el vector $outlier_{it}$ se modifica a $outlier_{it}^h$.

Criterio 3 – Anomalías en ventana temporal (Nuevo criterio)

En general los sistemas físicos poseen diferentes velocidades de respuesta o tiempos de reacción, es decir, que existe la posibilidad de medir eventos que suceden en instantes diferentes de tiempo pero que se hayan gatillado a partir del mismo fenómeno físico. Un ejemplo de lo anterior es: previo al momento donde se genera la disrupción sucede un fenómeno físico particular, el cual es medido a través de los diferentes sensores del dispositivo, puede darse el caso que, por ejemplo, la Corriente de Plasma (IP) genere una anomalía debido al fenómeno y que la señal Mode Lock también la genere, pero un lapso de tiempo después, en este caso estaríamos en presencia de tiempos de reacción de los sensores diferentes o el retraso es inherente al evento físico.

El anterior fenómeno no es considerado en los criterios de alarma de la etapa 1, por lo que la inclusión de nuevo criterio podría mejorar las tasas de acierto del sistema e incluso aumentar los tiempos de anticipación a la disrupción.

Descripción

Cuando no se haya gatillado alarma en ninguno de los dos primeros criterios, se realiza la evaluación de este nuevo criterio bajo la expresión matemática que se describe en (4-6).

$$Criterio_{3t} = \begin{cases} True, & \sum_{k=t+1-\delta}^t \sum_{\forall i} outlier_{it}^3 \geq \delta \\ False, & e. o. c \end{cases} \quad (4-6)$$

El criterio activa una alarma cuando la suma de todas las anomalías detectadas dentro de una ventana de tiempo de ancho δ (desde el instante $(t + 1 - \delta)$ hasta t) y para cualquier tipo de señal i supera un umbral también determinado por δ . En la Figura 4-12 se muestra de manera ilustrativa un caso donde se evalúa este nuevo criterio.

	Tiempo	Señales						
		IP	ML	PI	PD	SDE	RP	TIP
Suma Temporal	$t - 3$	1	0	0	0	0	0	0
	$t - 2$	1	0	0	0	0	0	0
	$t - 1$	0	1	0	0	0	0	0
	t	0	0	0	1	0	0	0
		Suma por Señal						

Figura 4-12: Ejemplo ilustrativo generación de alarma nuevo criterio

En el caso de la figura anterior, si δ se le asigna un valor de 4, se dispararía una alarma disruptiva. Cabe destacar que el parámetro asociado a este criterio solo puede tomar valores enteros positivos ($\delta \in \mathbb{N}$) y a medida que incrementa su valor se vuelve más exigente, disminuyendo la tasa de acierto y las falsas alarmas, en caso contrario aumentarían. Finalmente, para una mejor aclaración del funcionamiento de este criterio, se elaboró la Figura 4-13, la cual muestra el ejemplo anterior de manera gráfica.

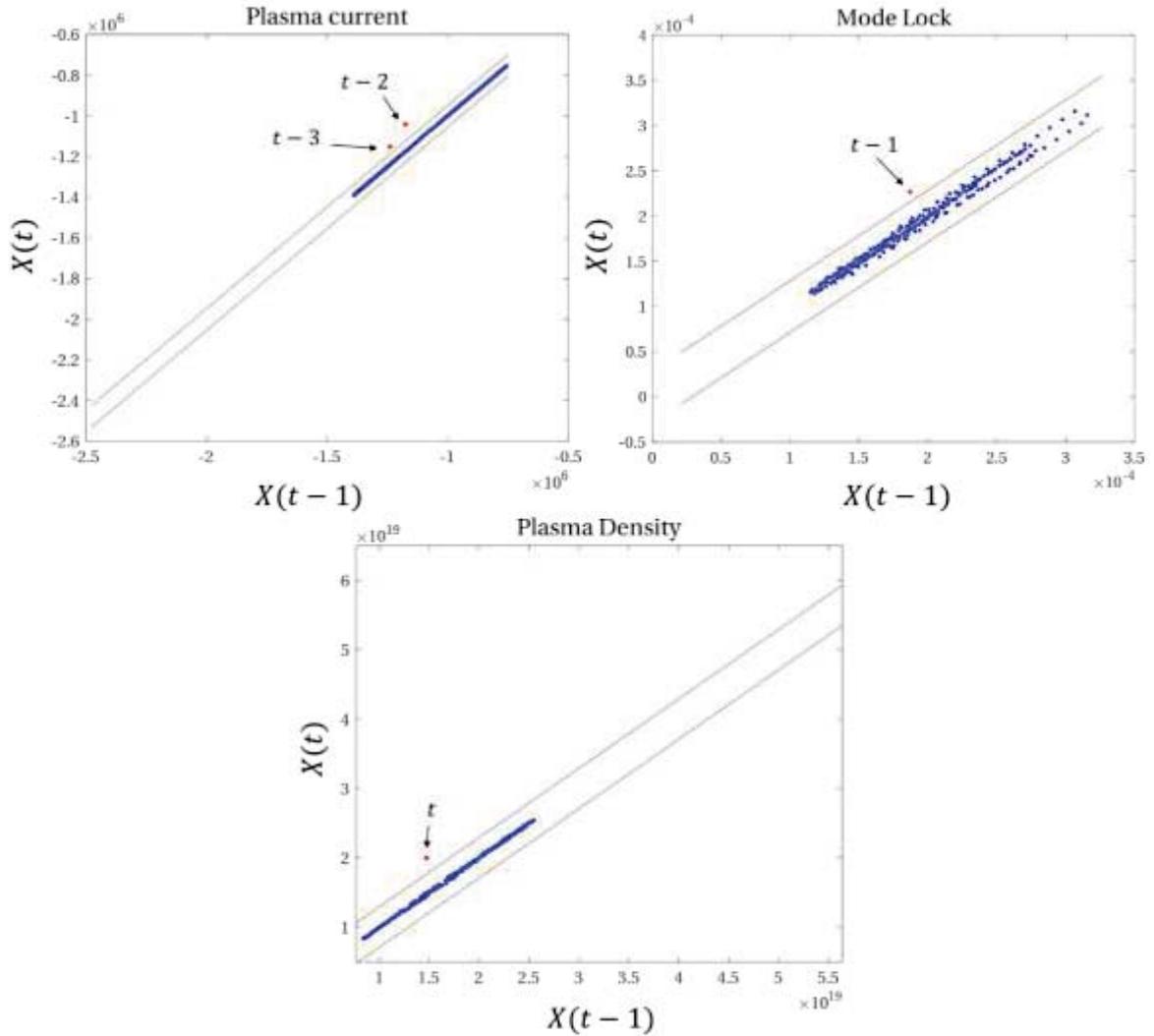


Figura 4-13: Ejemplo gráfico nuevo criterio de alarma

En el diagrama de bloques de la Figura 4-14 se muestra el sistema propuesto con las modificaciones comentadas en los apartados anteriores.

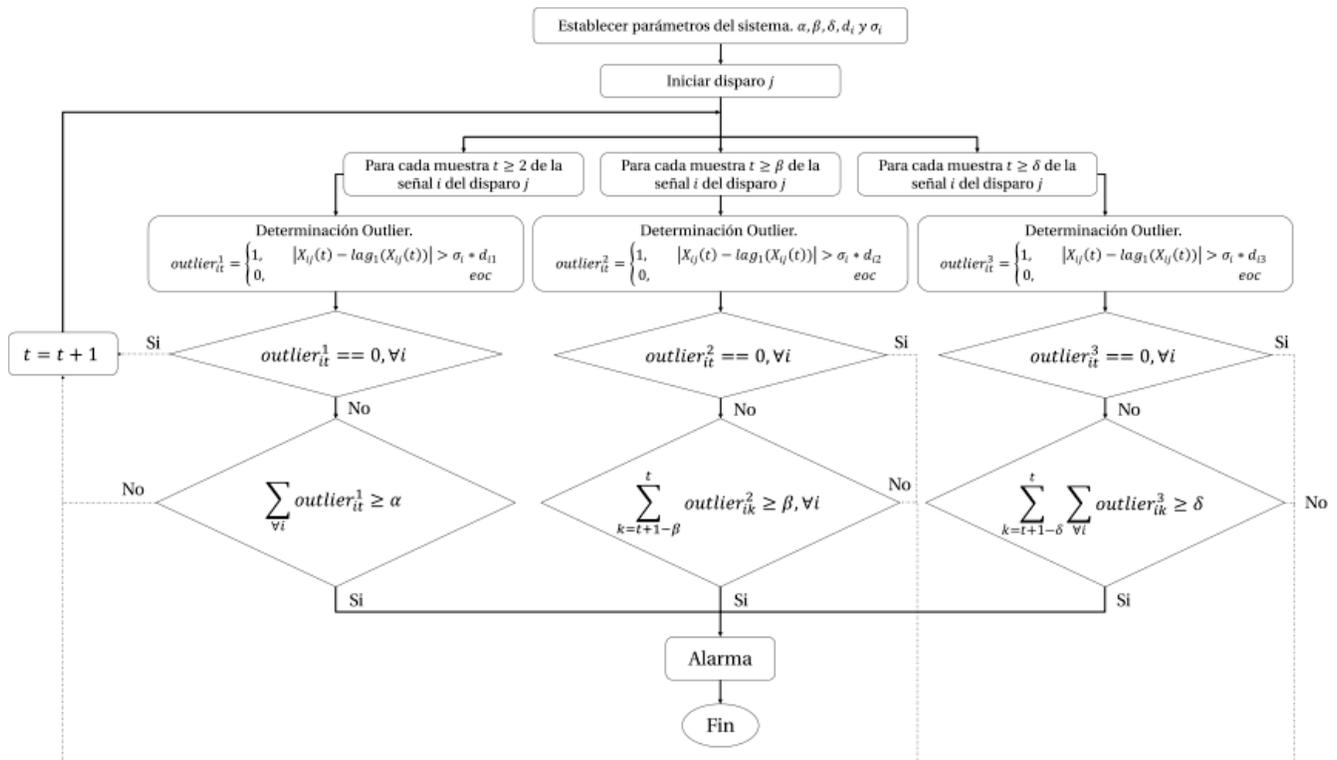


Figura 4-14: Diagrama en bloques del sistema de predicción propuesto en su segunda etapa

5 Entrenamiento DPAD

El proceso de entrenamiento del sistema DPAD se basa en el hecho de determinar el valor de sus parámetros de manera que se ajusten a los precursores que se generan antes de las interrupciones. Particularmente este sistema será abordado desde la perspectiva de un problema de optimización.

Un problema de optimización consiste en minimizar o maximizar el valor de una o varias funciones objetivo (FO). En otras palabras, se intenta determinar a qué valores de las variables que intervienen en la función objetivo, ésta se minimiza o maximiza. Es importante tener en cuenta las restricciones que afectan al problema, ya que éstas entregan el espacio de búsqueda donde es permisible una solución. Existen una gran cantidad de tipos de optimización, y dentro de cada uno, un sin número de métodos que se pueden adecuar de mejor manera de acuerdo al problema planteado. En el siguiente apartado se desarrolla brevemente el concepto de optimización metaheurística.

5.1 Optimización metaheurística

A diferencia de cualquier método de optimización clásico, los algoritmos metaheurísticos otorgan la posibilidad de solucionar problemas donde la dimensionalidad es alta, obteniendo soluciones más que aceptables en tiempos bastante más bajos que realizando una búsqueda exhaustiva. Es fundamental recalcar que las soluciones obtenidas mediante estos métodos no aseguran encontrar una solución óptima, sino que, una buena solución para el problema planteado. Su rapidez de funcionamiento y su capacidad de conseguir soluciones bastante cercanas al óptimo, ha generado un aumento en su uso en problemas de múltiples áreas, en [43] se detallan un conjunto de áreas donde estos métodos han logrado buenos resultados.

Existen un gran número de técnicas metaheurísticas en la literatura, cada una de ellas con sus propios parámetros y etapas que la conforman, entre estas destacan:

- Particle Swarm Optimization (PSO).
- Firefly Optimization.
- Bat Optimization Algorithm.
- Artificial Bee Colony (ABC).
- Genetics Algorithms (GA).

Dentro de las técnicas más populares en esta área se encuentran los Algoritmos Genéticos (GA), los cuales se describirán de forma resumida en el siguiente apartado.

5.1.1 Algoritmos genéticos

Los algoritmos genéticos son métodos de optimización generados a partir de la teoría de evolución de las especies de Darwin en 1859, esta teoría propone que, a lo largo de las generaciones la población de un de una especie específica evoluciona de acuerdo los principios de la selección natural y bajo el concepto de la supervivencia del más fuerte, además, incluye la noción de la similitud entre padres e hijos de acuerdo a ciertas características hereditarias (Genes). Lo anterior se extrapola a la optimización de una función objetivo específica, donde la población se conforma a partir de individuos que son evaluados en dicha FO para ver su respectiva adaptación al problema, para luego realizar la cruce de los individuos, la mutación y finalmente la selección. La cruce es la etapa que se encarga de componer las nuevas generaciones a partir de la combinación de los genes de los padres, dichos genes son los valores de las variables del problema. La mutación por su parte se encarga de alterar en algún grado específico los genes de los nuevos individuos obtenidos a partir de la cruce, finalmente se calculan sus respectivas aptitudes según la FO y se determina de acuerdo a esto último si sobreviven o no a la generación siguiente. Los principales aspectos a establecer en las etapas anteriormente comentadas son los siguientes:

- **Cruza:** En este proceso se debe establecer la probabilidad de cruce (P_{cruza}) entre dos individuos, además se debe optar por una forma de elección de los padres y el tipo de cruce que se va a utilizar.
- **Mutación:** Se debe determinar en este punto la probabilidad de mutación (P_{mut}) de los nuevos individuos y la manera en que mutan.
- **Selección o elitismo:** En esta etapa se debe establecer cómo será la selección entre los padres e hijos a partir de su aptitud para las siguientes generaciones.

Cabe destacar que los tipos de cruce y mutación (operadores genéticos) seleccionados están ampliamente relacionados con la codificación de los genes o variables del problema. En la Figura 5-1 se describe a partir de un diagrama de bloques el funcionamiento del GA de forma general.

5.2 Entrenamiento por conjunto

Típicamente los entrenamientos de un sistema clasificador se basan en la separación de dos conjuntos, uno de entrenamiento y otro de test, la idea es encontrar una regla lo más general posible a través del conjunto de entrenamiento y validarla a través del de test. De acuerdo a esto, a partir de la base de datos de 1035 descargas safe y 201 disruptivas, se dividieron de forma aleatoria los conjuntos de entrenamiento y test en 60(%) y 40(%) respectivamente (742 descargas de entrenamiento y 494 de test), lo cual corresponde a 621 descargas safe y 121 disruptivas para el entrenamiento y 414 safe más 80 disruptivas para el test.

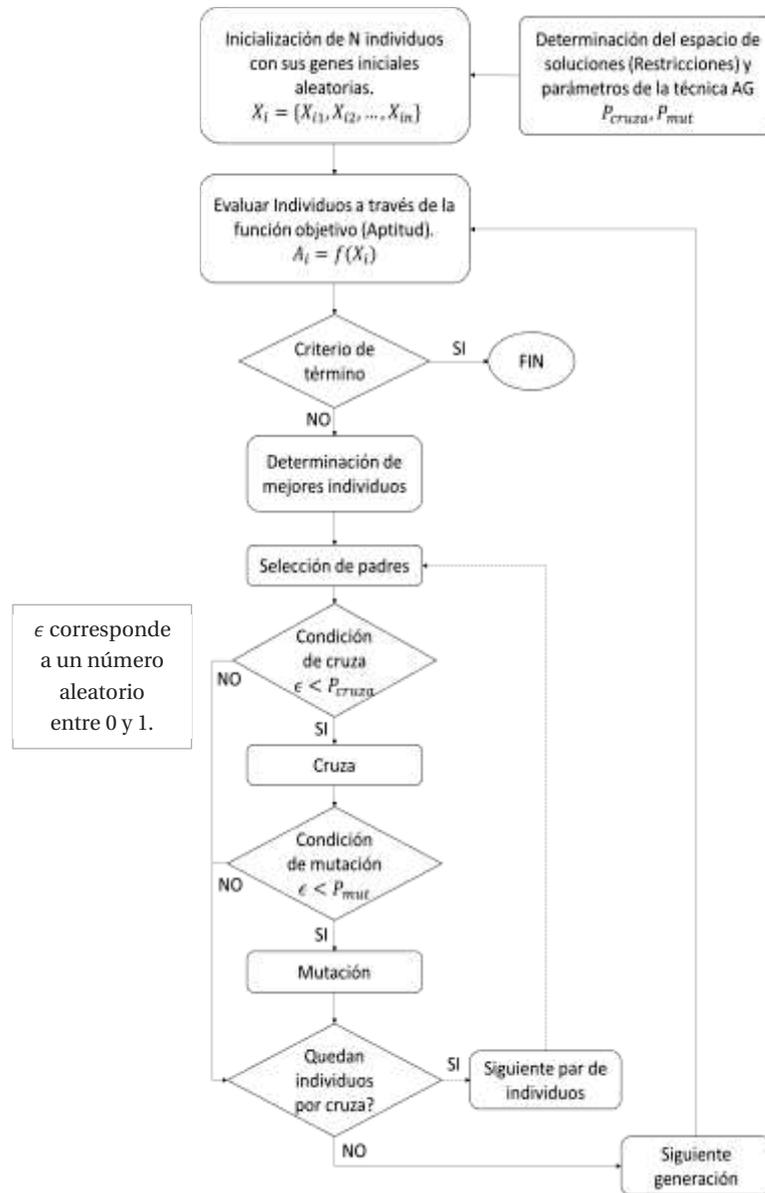


Figura 5-1: Diagrama en bloques del Algoritmo Genético

5.3 Implementación etapa 1

Mediante la implementación de la técnica de optimización GA se busca determinar una serie de parámetros para lograr que el sistema de predicción de disrupción logre generar una tasa de verdaderos positivos y falsos positivos lo más alta y baja posible respectivamente, para esto se estableció la función objetivo que se describe en (5-1).

$$FO = TPR * (1 - FPR) \quad (5-1)$$

Donde TPR corresponde a la tasa de verdaderos positivos y FPR al de los falsos positivos, los cuales a su vez dependen de 9 variables que se resumen en la Tabla 5-1 y donde se establecen sus respectivas restricciones.

Tabla 5-1 Variables de decisión y sus respectivas restricciones

Variable	Descripción	Restricción
d_i	Ajuste de umbral de anomalías para la señal de tipo i	$0 \leq d_i \leq 3, d_i \in \mathbb{R} \forall i$
α	Número de anomalías por señal mínimas para disparar alarma de interrupción (criterio 1)	$1 \leq \alpha \leq 7, \alpha \in \mathbb{N}$
β	Número de anomalías consecutivas en una misma señal mínimas para disparar alarma de interrupción (criterio 2)	$1 \leq \beta \leq 20, \beta \in \mathbb{N}$

Las restricciones de d_i y α se describen en la sección 4.3.1, para la variable β en cambio se acotó su espacio de búsqueda ya que para valores muy altos el criterio al cual pertenece se vuelve de una exigencia demasiado alta. En definitiva, el problema de optimización planteado se resume de acuerdo al siguiente planteamiento matemático.

$$\text{Max } FO = TPR * (1 - FPR) = f(d_i, \alpha, \beta), \forall i \quad (5-2)$$

S.A

$$0 \leq d_i \leq 3, d_i \in \mathbb{R} \forall i$$

$$1 \leq \alpha \leq 7, \alpha \in \mathbb{N}$$

$$1 \leq \beta \leq 200, \beta \in \mathbb{N}$$

Adicionalmente, el proceso de optimización implementado tiene 3 características importantes de destacar:

1. Selección elite: la cantidad de hijos catalogados como elite y que automáticamente avanzan a la siguiente iteración o generación es de 10, que equivale al 5(%) del tamaño de la población.
2. Fracción de cruce: el 80(%) de los hijos producidos por el cruce, sin contar a los de elite, pertenecen a la generación siguiente. Cabe destacar que la selección de los padres se realiza a través de la función ruleta, en la cual se le asigna una porción o probabilidad a cada individuo según su aptitud con la función objetivo, de manera que mientras mejor adaptado esté un individuo es más probable que se sea seleccionado, finalmente la cruce es mediante la selección al azar de los genes de dichos padres.
3. Mutación: se implementa la función de mutación gaussiano, la cual añade un número al azar tomado de una distribución gaussiana con media 0 y desviación estándar dada por los límites de los parámetros (por ejemplo para d_i su desviación estándar es de $3 - 1 = 2$) y además los factores de escala y encogimiento, lo anterior se describe de forma detallada en [44].

El desarrollo del proceso de optimización alcanzó un valor de la función objetivo de 0.9327, que corresponde a un TPR de 94.21(%) y FPR de 1.77(%). Los valores de los parámetros obtenidos se indican en la Tabla 5-2, y en la Figura 5-2 se observa también el proceso de optimización de los parámetros del modelo.

Tabla 5-2 Valor de los parámetros del modelo en la optimización

Parámetro	Valor
d_1 Corriente de Plasma	0,0984
d_2 Mode Lock	0,5674
d_3 Plasma internal inductance	2,9698
d_4 Plasma density	0,0651
d_5 Stored diamagnetic energy time derivative	1,4540
d_6 Radiated power	2,6042
d_7 Total input power	2,2591
α	2
β	5

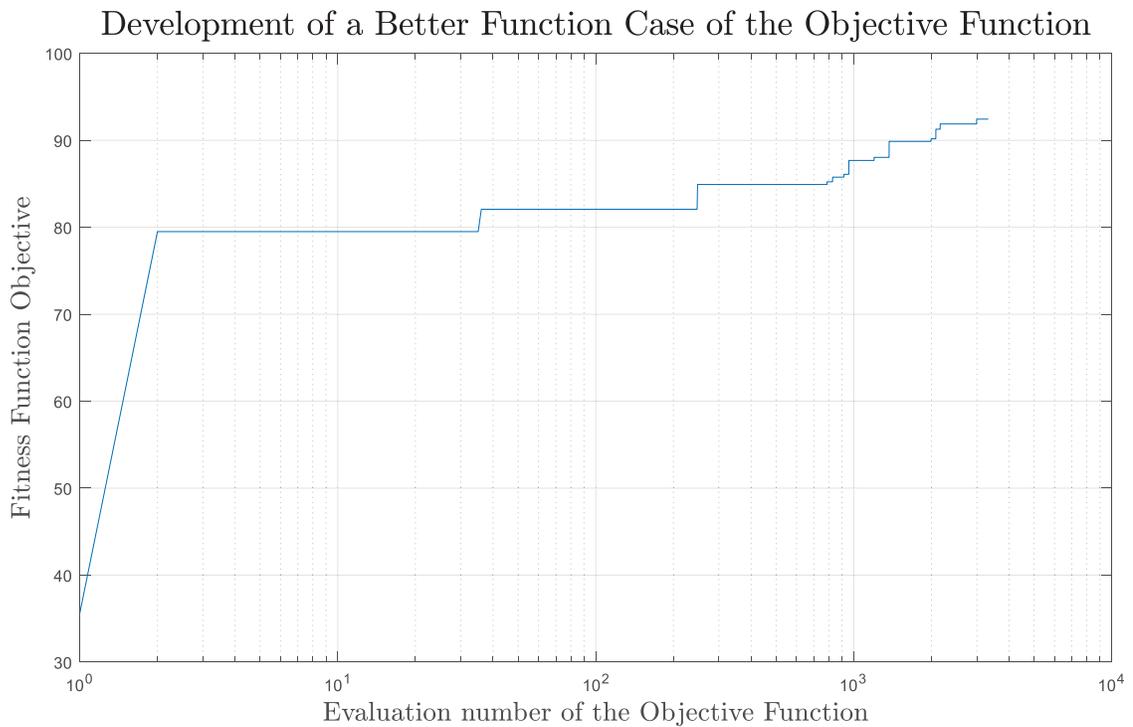


Figura 5-2: Desarrollo mejor caso aptitud FO por evaluación de la FO

5.4 Implementación etapa 2

La función objetivo y las restricciones de los parámetros α y β no se modifican, tan solo se altera la restricción asociada al ajuste de umbral (5-3) y se incluye la asociada al nuevo criterio (5-4).

$$0 \leq d_{ih} \leq 3, d_i \in \mathbb{R} \forall i, \forall h \quad (5-3)$$

$$1 \leq \delta \leq 20, \delta \in \mathbb{N} \quad (5-4)$$

Debido a que cada criterio funciona de manera paralela y el modo de entrenamiento implementado es a través de algoritmos genéticos, considerar un entrenamiento aislado por

critorio es una buena decisión debido principalmente a la reducción de la cantidad de parámetros a determinar (de 24 parámetros a 8 parámetros por optimización). Según lo anterior se obtuvieron los resultados de entrenamiento que muestran en la Tabla 5-3.

Tabla 5-3 Resultados entrenamiento por criterio

Criterio	<i>TPR</i> (%)	<i>FPR</i> (%)	<i>FO</i>
1	90.08	4.03	0.8645
2	89.26	2.58	0.8696
3	95.87	4.03	0.9201

Los parámetros obtenidos en cada entrenamiento se muestran en la Tabla 5-4 y el desarrollo de la función objetivo a través de las diferentes generaciones del algoritmo se muestra en la Figura 5-3.

Tabla 5-4 Valor de parámetros obtenidos a través del entrenamiento por criterio

Parámetro	Criterio 1	Criterio 2	Criterio 3
d_{1h} Corriente de Plasma	1.8594	1.2713	2.3988
d_{2h} Mode Lock	1.4610	0.5667	0.4529
d_{3h} Plasma internal inductance	0.0061	2.8664	2.6087
d_{4h} Plasma density	0.5239	0.1889	0.0209
d_{5h} Stored diamagnetic energy time derivative	2.8621	1.4637	0.8431
d_{6h} Radiated power	2.7666	1.6050	1.0046
d_{7h} Total input power	2.4962	2.6962	1.7888
α	2	-	-
β	-	5	-
δ	-	-	7

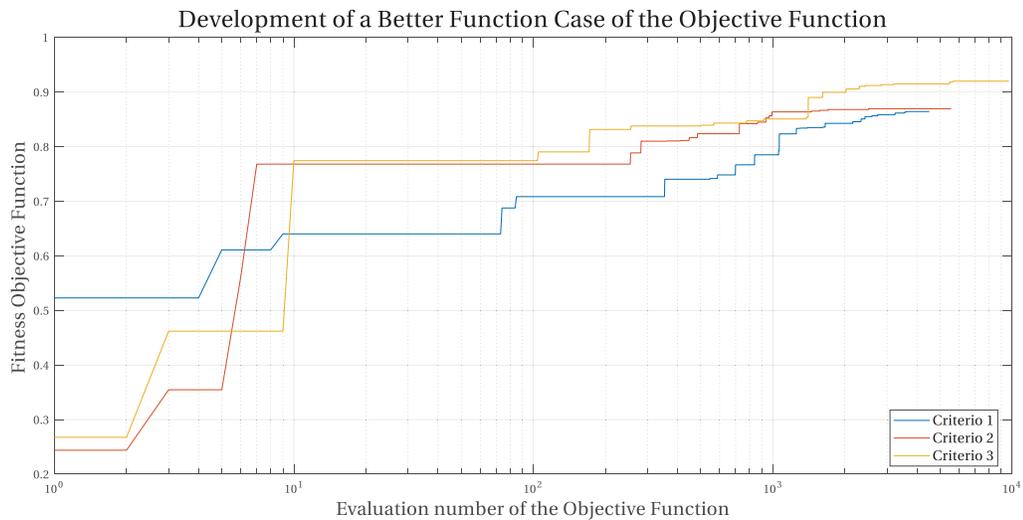


Figura 5-3: Desarrollo mejor caso aptitud FO por evaluación de la FO para cada criterio

5.4.1 Entrenamiento de refinamiento

A pesar que los criterios pueden trabajar de forma aislada, no necesariamente el óptimo individual de cada uno corresponde al óptimo general del sistema, pero es apropiado suponer que este óptimo global esté relativamente cercano a los valores de parámetros que se lograron previamente. En relación a lo expuesto se realizó nuevo entrenamiento de refinamiento, el cual intenta conseguir que los criterios se complementen adecuadamente y en definitiva se logre una mayor cercanía al óptimo general del predictor. Finalmente, en esta última fase de entrenamiento se lograron las tasas que se muestran en la Tabla 5-5, para los casos con y sin refinamiento, además en la Figura 5-4 se describe la variación de la función objetivo a través de las evaluaciones realizadas en ella. Es importante señalar que la unión de los criterios se realiza a partir del operador OR, es decir que la alarma disruptiva se lanza cuando al menos uno de los criterios se hace verdadero.

Tabla 5-5 Resultados entrenamiento sistema completo con y sin refinamiento

Entrenamiento	<i>TPR</i> (%)	<i>FPR</i> (%)	<i>FO</i>
Sin Refinamiento	97.52	7.41	0.903
Con Refinamiento	96.69	6.12	0.9078

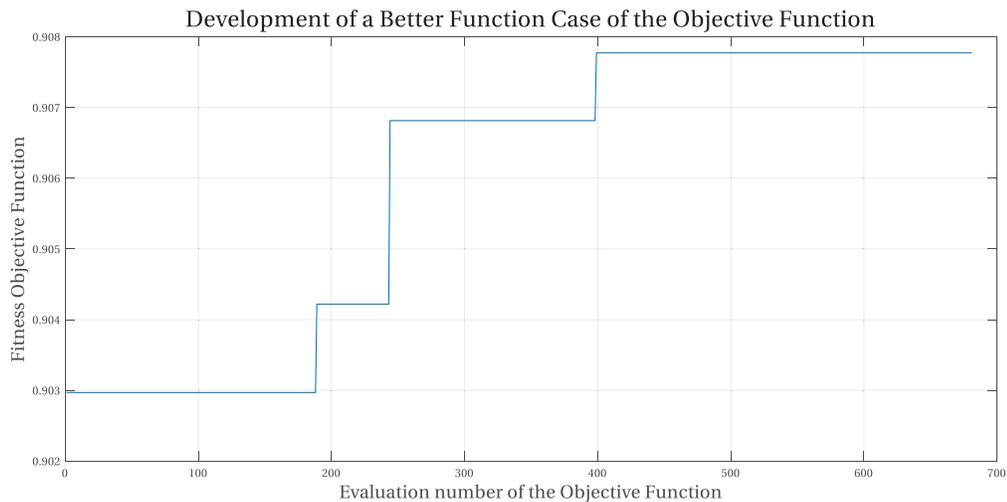


Figura 5-4: Desarrollo mejor caso aptitud FO para el entrenamiento de refinamiento

Como se observa en la Tabla 5-5 y Figura 5-4, la función objetivo mejoró en tan solo 0.0048 lo cual corresponde en una disminución en la *TPR* del 1.27(%) y la *FPR* del 1.29(%). Estos valores indican que la variación no es significativa, pero es necesario evaluar cada conjunto de parámetros mediante el test para realizar una comparación adecuada. Adicionalmente, en la Tabla 5-6 se muestran los valores de los parámetros luego del entrenamiento de refinamiento.

Tabla 5-6 Valor de parámetros obtenidos a través del entrenamiento de refinamiento

Parámetro	Criterio 1	Criterio 2	Criterio 3
d_{1h} Corriente de Plasma	1,8593	1,2712	2,3987
d_{2h} Mode Lock	1,4609	0,5666	0,4528
d_{3h} Plasma internal inductance	0,0061	2,8664	2,6086
d_{4h} Plasma density	0,5239	0,1889	0,0208
d_{5h} Stored diamagnetic energy time derivative	2,8621	1,4636	0,8431
d_{6h} Radiated power	2,7666	1,6050	1,0045
d_{7h} Total input power	2,4962	2,6962	1,7888
α	2	-	-
β	-	6	
δ	-	-	9

De acuerdo a la comparación de la Tabla anterior y la Tabla 5-4 se observan variaciones significativas en los parámetros β y δ , para el resto el efecto es prácticamente nulo. También de la tabla anterior se puede extraer información respecto a las señales más sensibles para detectar anomalías y por ende aquellas que pudiesen gatillar un mayor número de alarmas disruptivas. La sensibilidad está estrechamente relacionada con el ajuste de umbral, entonces mientras más pequeño sea este valor, más propenso es el sistema para detectar outliers, conforme a lo anterior se realizó la Tabla 5-7 donde se indican por criterio las tres señales con mayor sensibilidad de detección.

Tabla 5-7 Tipos de señales con ajuste de umbral más bajo para cada criterio

Criterio 1	Criterio 2	Criterio 3
Plasma internal inductance	Plasma density	Plasma density
Plasma density	Mode Lock	Mode Lock
Mode Lock	Corriente de Plasma	Stored diamagnetic energy time derivative

Como finalización de este capítulo, es correcto indicar que los procesos de entrenamiento realizados mediante algoritmos genéticos muestran que fue posible lograr una convergencia, de manera que existe un conjunto de parámetros que logran a través del sistema DPAD en su primera y segunda etapa detectar precursores disruptivos, no obstante, es necesario verificar los resultados a través del conjunto de test y así demostrar la generalidad del predictor.

6 Resultados

En este capítulo se mostrarán los resultados obtenidos tanto para las tasas de acierto y los tiempos de anticipación de las etapas del sistema DPAD, a continuación, se realiza una comparativa entre diferentes predictores. Finalmente, se presenta un análisis del comportamiento del sistema según las señales más importantes y las curvas ROC de sus parámetros.

6.1 Etapa 1

En el entrenamiento efectuado se utilizó el 60(%) de los disparos, correspondientes a 742 descargas (621 safe y 121 disruptivas), para la validación del modelo obtenido se evaluó el conjunto de test (494 descargas, equivalente al 40(%) del total), obteniendo tasas de verdaderos positivos y falsos positivos de 90(%) y 2.66(%) respectivamente, es decir que de las 80 descargas disruptivas que forman parte del conjunto de test, 72 fueron bien predichas y de los 414 disparos safe, en 11 se gatilló la alarma disruptiva. Por otro lado, respecto a los tiempos de anticipación de la disrupción (WarningTime), se obtuvo una media de 430.5(ms) con una desviación estándar de 581.1(ms), en la Figura 6-1 se muestra detalladamente lo anterior a través de la fracción acumulativa de las interrupciones detectadas a diferentes WarningTime.

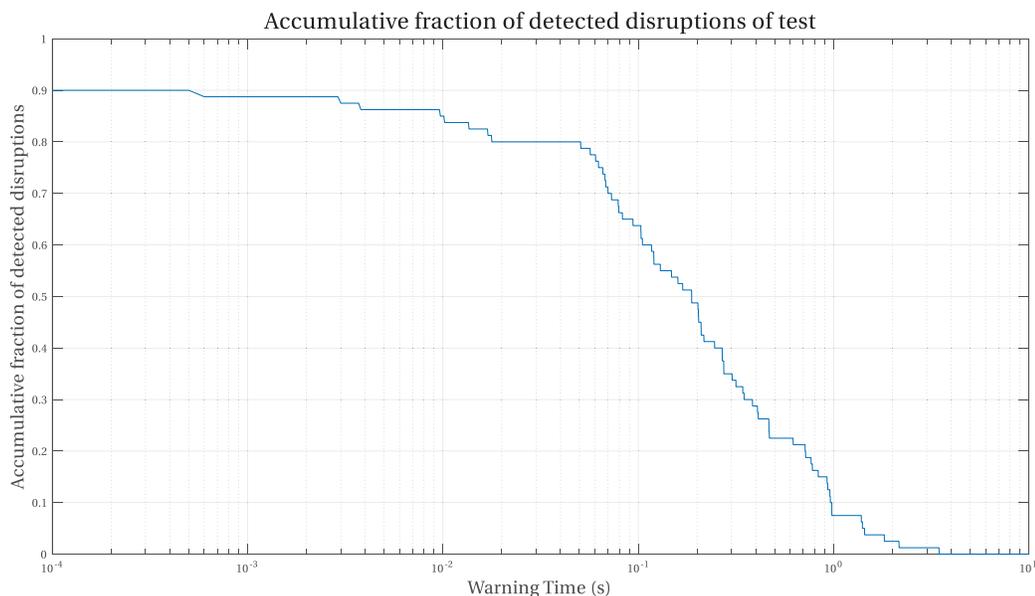


Figura 6-1: Fracción acumulativa de las interrupciones detectadas mediante el primer sistema DPAD

La manera de interpretar la curva anterior es la siguiente: un par WT_1 (eje de abscisas) y fracción de detección (eje de las ordenadas) indica la relación de disparos detectados con un tiempo de anticipación mayor a WT_1 respecto al total de disparos disruptivos.

Adicionalmente, en [36] se establecen diferentes parámetros para la evaluación del tiempo de anticipación a la disrupción, los cuales se describen y se muestra lo obtenido en la Tabla 6-1.

Tabla 6-1 Resultados obtenidos del primer sistema predictor propuesto según [36]

Parámetro	Descripción	Expresión Matemática	Valor
False Alarm (FA)	Alarma gatillada en una descarga no-disruptiva.	-	2.66(%)
Missed Alarm (MA)	Alarma no se gatilla en una señal disruptiva.	-	1.25(%)
Premature Alarm (PA)	Alarma gatillada en tiempos mayores a 1(s) antes de la disrupción.	$WT > 1(s)$	7.5(%)
Tardy Detection (TD)	Alarma gatillada en tiempos menores de 30(ms) en una señal disruptiva.	$WT < 30(ms)$	18.75(%)
Valid Alarm (VA)	Alarma gatillada entre 30(ms) y 1(s) en una señal disruptiva.	$30(ms) < WT < 1(s)$	72.5(%)
Success Rate (SR)	Total de predicciones correctas.	$VA + PA + TD$	98.75(%)
Alarmas Positivas	Alarmas gatilladas antes de la disrupción	$WT > 0(s)$	90(%)
Acierto Total	Todas las alarmas bien predichas	-	96.15(%)

En general los resultados obtenidos son bastante buenos a pesar de ser un sistema predictor tan sencillo, no obstante, es necesario realizar mejoras para lograr los requerimientos establecidos por ITER. Según la premisa anterior se realizaron una serie de modificación (sección 4.3.3), obteniendo los resultados que se muestran en el siguiente apartado.

6.2 Etapa 2

6.2.1 Resultados entrenamiento por criterio

Los resultados obtenidos a partir de la evaluación del conjunto de test para cada set de parámetros entrenados de manera separada y la unión de ellos mediante el operador lógico OR se presentan en la Tabla 6-2, de la cual se extrae que existe un grado de complementación en los parámetros, en consecuencia, se mejora en cierto grado la tasa de acierto, pero se afecta a la tasa de falsos positivos. Acorde a esto y para corroborar la aseveración anterior se elaboró la Tabla 6-3, que muestra la participación de los criterios en el disparo de alarmas disruptivas (independiente del warning time obtenido).

Tabla 6-2 Resultado por criterio y sistema completo (sin refinamiento) mediante conjunto de test

Criterio	TPR(%)	FPR(%)
1	93.75	4.59
2	85	2.9
3	88.75	4.59
1, 2 y 3	96.25	8.45

Tabla 6-3 Participación de criterios del sistema completo (sin refinamiento)

Criterio	Número de Alarmas		Porcentaje del Total de alarmas
	Correctas	Incorrectas	
1	42	16	52.5(%)
2	10	7	12.5(%)
3	28	12	35(%)
1, 2 y 3	80	35	100(%)

El criterio de anomalías simultáneas para este conjunto de parámetros (sin refinamiento) es el que posee un mayor nivel de participación, pero las diferencias son suficientemente aptas para aseverar que existe un complemento entre criterios. Adicionalmente, de la tabla anterior se observa la importancia de reducir la cantidad de alarmas incorrectas, es bajo este punto donde se hace fundamental realizar un entrenamiento de refinamiento.

Finalmente es importante destacar que no se indican todos los parámetros y análisis para este primer entrenamiento, sino que los datos entregados anteriormente son fundamentalmente para reafirmar la necesidad de un entrenamiento de refinamiento, el cual busca evitar el hecho de que un criterio se puede detectar una interrupción específica, pero que esto a su vez genera que dispare alarmas falsas a pesar de que otro criterio lo realiza sin la necesidad de afectar la FPR, es decir que se busca disminuir la sensibilidad de un criterio debido a que otro ya fue capaz de detectar la interrupción sin perjudicar los resultados del predictor.

6.2.2 Resultado entrenamiento de refinamiento

En este último caso se obtuvieron tasas de verdaderos positivos y falsos positivos de 96.25(%) y 6.52(%) respectivamente, es decir que de las 80 descargas disruptivas que forman parte del conjunto de test, 77 fueron bien predichas y de los 414 disparos safe, en 27 se gatilló la alarma disruptiva. En definitiva, el refinamiento fue capaz de disminuir en más de 2 puntos porcentuales la tasa de falsos positivos y mantener la TPR. Como resumen se muestra en la Tabla 6-4 los resultados del sistema DPAD en su segunda etapa con y sin entrenamiento de refinamiento, además del individual de cada criterio.

Tabla 6-4 Resumen de resultados del sistema DPAD en su segunda etapa

	<i>TPR</i> (%)	<i>FPR</i> (%)	<i>FO</i>
DPAD (sin refinamiento)	96.25	8.45	0.8812
Criterio 1	93.75	4.59	0.8945
Criterio 2	85	2.9	0.8254
Criterio 3	88.75	4.59	0.8468
DPAD (con refinamiento)	96.25	6.52	0.8997
Criterio 1	93.75	4.59	0.8945
Criterio 2	80	1.93	0.7845
Criterio 3	83.75	2.90	0.8132

La tabla anterior muestra como los criterios 2 y 3 disminuyeron su sensibilidad, lo cual se ve traducido en una disminución de la *TPR* y la *FPR*, generando entonces que estos criterios tengan una peor aptitud, pero mirado desde el punto de vista global del predictor, si existe una mejora en las tasas de evaluación, lográndose por ende el objetivo del entrenamiento de refinamiento.

Desde el punto de vista temporal, los tiempos de *WarningTime* alcanzaron una media de 830(*ms*) y una desviación estándar de 1.26(*s*), en la gráfica de la Figura 6-2 se muestra la fracción acumulativa de las interrupciones detectadas a diferentes tiempos de anticipación para la segunda etapa del sistema propuesto.

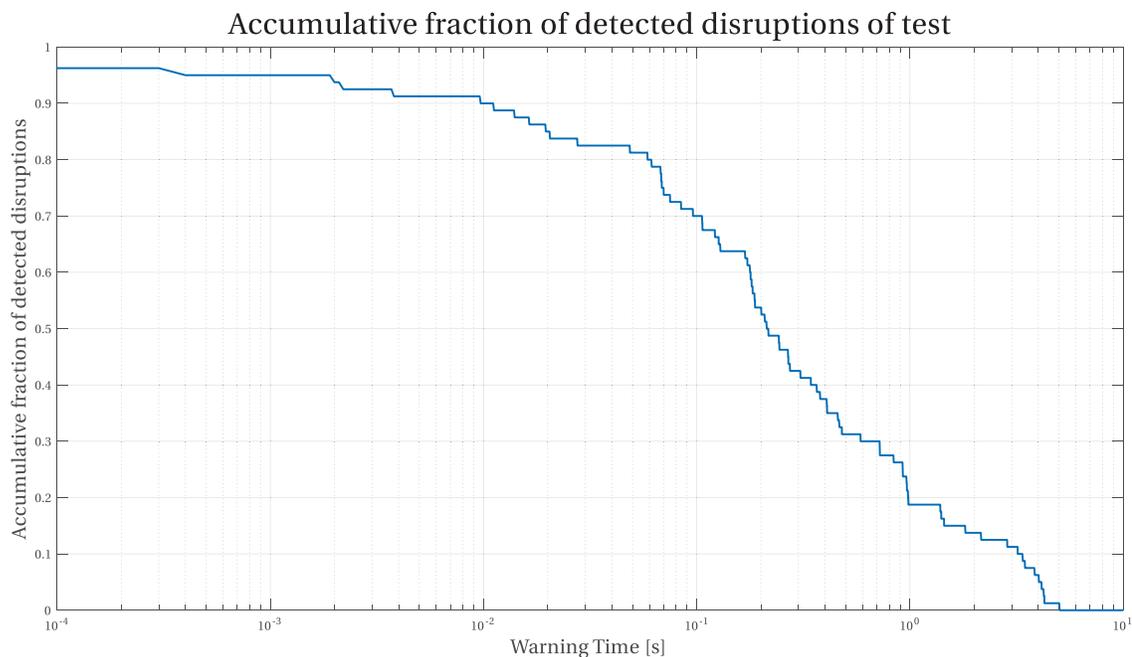


Figura 6-2: Fracción acumulativa de las interrupciones detectadas mediante el segundo sistema DPAD

Junto a la gráfica anterior se construye la Tabla 6-5, la cual muestra los resultados obtenidos a partir de una serie de parámetros (al igual que los resultados de la primera etapa del sistema DPAD).

Tabla 6-5 Resultados obtenidos del segundo sistema predictor propuesto según [36]

Parámetro	Expresión Matemática	Valor
False Alarm (FA)	-	6.52(%)
Missed Alarm (MA)	-	0(%)
Premature Alarm (PA)	$WT > 1(s)$	18.75(%)
Tardy Detection (TD)	$WT < 30(ms)$	17.5(%)
Valid Alarm (VA)	$30(ms) < WT < 1(s)$	63.75(%)
Success Rate (SR)	$VA + PA + TD$	100(%)
Alarmas Positivas	$WT > 0(s)$	96.25(%)
Acierto Total	-	93.93(%)

Al igual que el sistema predictor en su etapa 1, no se logran los requerimientos propuestos por ITER, pero en este caso los resultados están muy cercanos, teniendo por un lado que la tasa de verdaderos positivos supera el umbral del 95(%) en 1.25(%) pero la de falsos está por encima del 5(%) en 1.52(%). Para un correcto análisis de los resultados obtenidos, es sustancial realizar la comparación con modelos anteriores tales como APODIS.

6.3 APODIS

APODIS [32] (The Advance Predictor Of Disruptions, predictor avanzado de interrupciones en español), este proyecto comenzó en el año 2008 con el objetivo de desarrollar un predictor de interrupciones para el dispositivo JET mejor que los anteriores. Las señales utilizadas por APODIS para identificar un comportamiento disruptivo son más o menos las mismas que las utilizadas en los predictores anteriores y las utilizadas en el predictor propuesto. Sin embargo, APODIS proporciona una novedad importante en relación con otros predictores de interrupción. Sigue una arquitectura de varios niveles (Figura 6-3) en la que los tres clasificadores de primer nivel se combinan con otro clasificador en el segundo nivel.

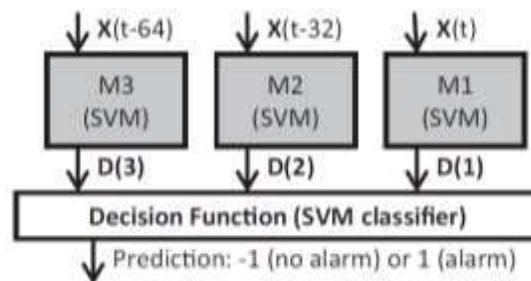


Figura 6-3: Arquitectura APODIS [32]

Cada vector de 32 muestras pasa por un proceso de extracción de características para luego ser las entradas de la primera capa de SVM (Máquina de Vectores Soporte), las distancias respecto al hiperplano en cada caso son a su vez las entradas del SVM clasificador. En definitiva, APODIS ha alcanzado resultados superiores en comparación a un sinnúmero de otros predictores, por lo que se hace interesante compararlo con los resultados del sistema propuesto.

6.4 Comparativa

Un elemento importante a resaltar es la comparación de los resultados del predictor respecto a modelos como APODIS, el que se presenta en la sección 3.5 y las dos versiones implementadas, con la finalidad de tener un mejor conocimiento del comportamiento del sistema. Según lo anterior, en la Tabla 6-6 se resumen los resultados a modo de comparación entre los diferentes predictores comentados.

Tabla 6-6 Comparativa de predictores

	Tasa de Acierto (%)	Alarmas Perdidas (%)	Falsas Alarmas (%)	Warning Time promedio (ms)	Tasa de Acierto (%) (Warning Time >30(ms))
APODIS	98.36	1.64	0.92	426	87.5
Basado en AD [31]	89.4	10.6	8.98	389	80
DPAD 1	90	10	2.66	430.5	80
DPAD 2	96.25	3.75	6.52	830	84

- DPAD 1 versus DPAD 2: La diferencia en TPR es de 6.25(%) a favor del predictor con las modificaciones realizadas y para el FPR el modelo anterior es inferior en 3.86(%), respecto a la función objetivo hubo un aumento desde 0.8791 a 0.8997. De acuerdo a lo anterior se establece que las mejoras del sistema fueron capaces de lograr mejores resultados.
- DPAD 2 versus Basado en AD: Las tasas de medición de los modelos propuestos (etapa 1 y 2) superan los resultados que entrega el predictor estudiado, logrando una TPR superior en 6.85(%) y una FPR inferior en 2.46(%). Lo que lleva a aseverar que este enfoque propuesto a pesar de su sencillez y la falta de una etapa de extracción de características, logra resultados superiores que sistemas más complejos.
- APODIS versus DPAD 2: El contraste entre los resultados obtenidos para estos dos métodos de predicción indican que; respecto a la tasa de verdaderos positivos los valores son bastante cercanos (diferencia a favor de APODIS de 2.11(%)); por otro lado la tasa de falsos positivos que alcanza APODIS es menor al predictor propuesto en 6.52(%), lo anterior expuesto le quita al sistema un grado de confianza a la hora de gatillar una alarma disruptiva.

Desde el punto de vista temporal también es interesante realizar la comparación de cada modelo, de acuerdo a esto se muestra en la Figura 6-4 la fracción acumulativa de las interrupciones detectadas a diferentes WarningTime para cada uno de los predictores. Respecto a la media del WT (AWT) se observa que el predictor basada en AD, junto a APODIS y DPAD 1 tienen valores relativamente cercanos en comparación a DPAD 2, el cual es superior por cerca de 400(ms), esto se refleja en la gráfica, en donde se observa que para tiempos de anticipación mayores a 30(ms) la curva negra está por encima de cualquiera de las otras. Asimismo, el aumento del AWT tiene su origen bajo el hecho de que el predictor DPAD 2 tienen una gran cantidad de disparos disruptivos

en tiempos altos, lo cual genera en definitiva un aumento en la dispersión del tiempo de anticipación y lo cual deriva finalmente, en la problemática de predecir la diferencia de tiempo entre el momento que se dispara la alarma y cuando sucede la interrupción. Este último punto es planteado como trabajos futuros del proyecto.

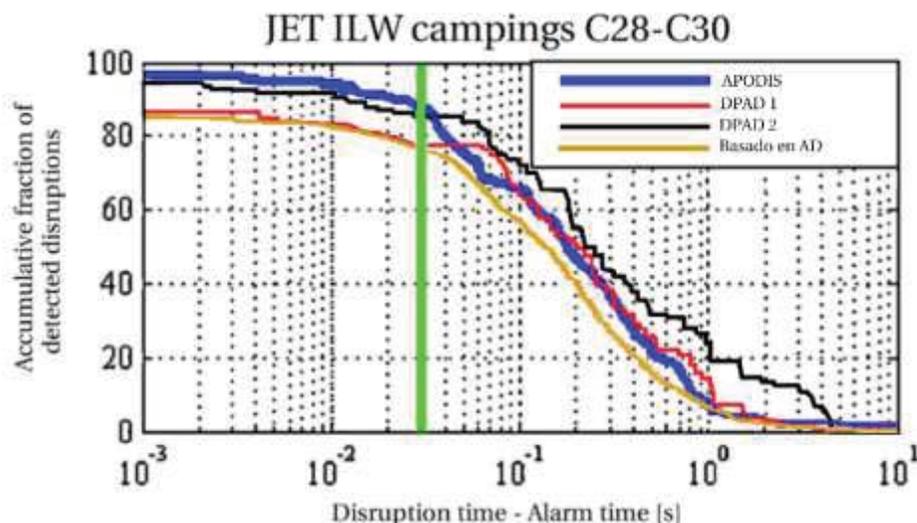


Figura 6-4: Comparación de fracción acumulativa de las interrupciones detectadas

6.5 Análisis DPAD

Para contestar la pregunta de qué señales dentro de las seleccionadas participan en una mayor cuantía dentro del sistema de predicción de interrupciones (DPAD 2), se construyó la Tabla 6-7, donde se indica la cantidad de participaciones y su porcentaje asociado.

Tabla 6-7 Participación de señales en DPAD 2

Nombre señal	Número de participaciones en alarmas	Porcentaje de participaciones en alarmas
Plasma current	0	0(%)
Mode Lock	165	47.14(%)
Plasma internal inductance	55	15.71(%)
Plasma density	92	26.29(%)
Diamagnetic energy derivative	4	1.14(%)
Radiated power	34	9.7(%)
Total input power	0	0(%)

Ahora bien, las señales que entregaron mayor información a la hora de gatillar alarmas por criterio, que señales generaron mayor número de anomalías para gatillar éstas y la cantidad de outlier totales detectados son resumidos en la Tabla 6-8.

Tabla 6-8 Participación de señales por criterio en DPAD 2. PND: Porcentaje de anomalías detectadas. NAD: Número de anomalías detectadas (generadoras de alarma). NATD: Número de anomalías totales detectadas.

Señal	Criterio 1			Criterio 2			Criterio 3		
	PAD(%)	NAD	NATD	PAD(%)	NAD	NATD	PAD(%)	NAD	NATD
1	0	0	0	0	0	0	0	0	0
2	38	39	50	86	43	2342	41.7	90	4956
3	50	51	621359	8	4	15	4.2	9	17
4	3.6	5	14	6	3	64	39	86	648
5	0	0	0	0	0	2	1.8	4	1117
6	8.3	7	7	0	0	842	12.5	27	4081
7	0	0	0	0	0	0	0	0	4

Para una visualización más simple, se elaboró la Tabla 6-9 de las señales más importantes por criterio y del sistema predictor en general.

Tabla 6-9 Ranking de participación de señales por criterio y sistema completo

Ranking Participación	Criterio 1	Criterio 2	Criterio 3	Total Criterios
1	Plasma internal inductance	Mode Lock	Mode Lock	Mode Lock
2	Mode Lock	Plasma internal inductance	Plasma density	Plasma density
3	Radiated power	Plasma density	Radiated power	Plasma internal inductance

En resumen, son tres las señales que intervienen mayormente en cada uno de los criterios y en el predictor total, donde destaca en gran medida la señal Mode Lock, la cual, y como se ha comentado en diferentes artículos, muestra muy claramente precursores disruptivos, pero a tiempos muy cercanos al fenómeno, lo anterior conlleva a concluir que el sistema es apto para trabajar en mitigación, pero no para evitación, ya que para esto último se requieren WT significativamente más grandes.

Además del análisis anterior, es importante observar si los criterios tienen un grado de complementación, para esto se evaluó el sistema completo (funcionando con los tres criterios a través del operador lógico OR) y se contaron la cantidad de participaciones. Cabe destacar que se le asigna una participación a un criterio cuando éste es el primero en disparar la alarma disruptiva.

Tabla 6-10 Participación de criterios del sistema completo DPAD 2.

Criterio	Número de Alarmas Correctas	Número de Alarmas Incorrectas	Porcentaje del Total de alarmas
1	51	18	63.75(%)
2	5	3	6.25(%)
3	24	6	30(%)
1,2 y 3	80	27	100(%)

Los resultados obtenidos denotan que existe complementación más que aceptable entre el criterio 1 y los dos restantes, ahora bien, si se comparan todos los criterios la participación del segundo es muy baja, esto se debe principalmente a la similitud de funcionamiento entre los criterios 2 y 3, donde ambos se enfocan en anomalías que se generan de forma consecutiva, con la diferencia que uno se aplica por señal y otro es independiente a esto.

En este primer estudio del sistema se utilizó para agrupar los criterios el operador lógico OR, entonces un análisis interesante es ver los resultados obtenidos si se modificada este operador o se realizan combinaciones de estos. En razón de lo señalado se construyó la Tabla 6-11.

Tabla 6-11 Combinación de criterios. C_1 , C_2 y C_3 son los criterios 1, 2 y 3 respectivamente

Combinación	TPR(%)	FPR(%)
$C_1 \wedge C_2 \wedge C_3$	75	1.21
$(C_1 \vee C_2) \wedge C_3$	82.5	1.69
$C_1 \wedge (C_2 \vee C_3)$	85	1.21
$(C_1 \vee C_3) \wedge C_2$	80	1.69
$(C_1 \wedge C_2) \vee C_3$	87.5	2.90
$(C_1 \wedge C_3) \vee C_2$	86.25	1.93
$(C_2 \wedge C_3) \vee C_1$	76.25	1.69
$C_1 \vee C_2 \vee C_3$	96.25	6.25

De igual forma como se esperaba, las diferentes combinaciones realizadas generaron un mejor en la tasa de falsos positivos, pero perjudico en gran medida la tasa de verdaderos positivos. En definitiva, la combinación de los criterios que mejor eficacia posee y se acerca de mejor manera a los requerimientos de ITER es el que se realiza mediante el operador OR.

Como se ha comentado a lo largo de este trabajo, el predictor propuesto tiene como una de sus principales cualidades el hecho de comportarse como caja transparente, lo cual entrega la posibilidad de modificar parámetros y anticiparse a los resultados obtenidos, es así como si se aumenta la sensibilidad de un umbral de detección este generaría el aumento en la TPR y FPR. Cabe destacar que este aspecto no lo posee APODIS. En virtud de lo expuesto y para complementar el análisis de sistema DPAD, es interesante construir las curvas ROC asociadas a cada parámetro.

6.5.1 Curva ROC

La curva ROC o Receiver operating characteristic (Característica Operativa del Receptor, en español) es aplicable tan solo a problemas de clasificación donde existen dos clases diferentes a discriminar. Consiste en una descripción gráfica que ilustra el rendimiento de un sistema de clasificación a medida que algún parámetro de configuración de éste es modificado. Las medidas de rendimiento que se suelen utilizar son la sensibilidad, que corresponde a la tasa de verdaderos positivos y la especificidad, la cual viene dada por resta de la unidad con la tasa de falsos positivos (también es común utilizar directamente la FPR). Un ejemplo de curva ROC se muestra en Figura 6-5, donde es posible observar el “trade-off” existente entre la tasa de verdaderos positivos y la tasa de falsos, es decir en nuestro caso, cuantas alarmas disruptivas bien realizadas gana al permitir equivocarme en disparos safe, o en su contraparte, cuantas alarmas pierdo al disminuir los falsos positivos. El clasificador perfecto se encuentra en el extremo superior izquierdo, donde se obtiene un 100(%) de TPR y 0(%) de FPR, pero que en la práctica es muy difícil que ocurra.

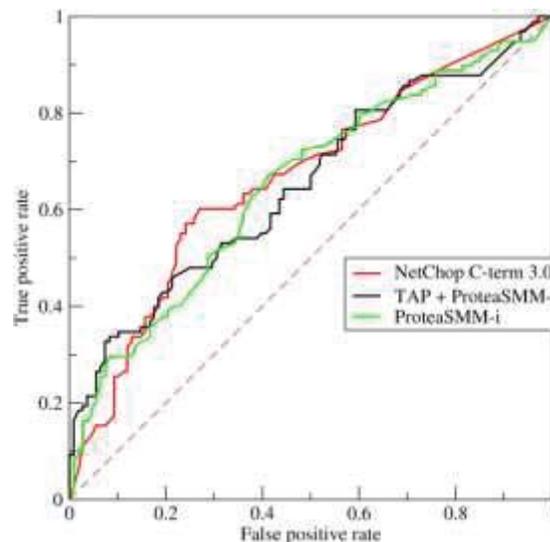


Figura 6-5: Ejemplo curva ROC (fuente: <https://en.wikipedia.org>)

En particular para el predictor DPAD, las curvas ROC construidas se realizan a partir del barrido de los 24 parámetros que interfieren, todos los ajustes de umbral d_{ih} se variaron desde el 10(%) del valor obtenido del entrenamiento hasta el 190(%). Para los parámetros α , β y δ tomaron los valores [1, 2, 3, 4], [3, 4, 5, 6, 7, 8, 9, 10] y [6, 7, 8, 9, 10, 11, 12] respectivamente.

Curva ROC – Criterio 1

La Figura 6-6 describe el comportamiento del sistema para la modificación de los parámetros de ajuste de umbral que más alteraron la TPR y FPR junto al parámetro asociado al criterio. El traslado en las gráficas de derecha a izquierda sobre las curvas corresponde a un aumento en el parámetro, entonces para el mínimo valor de éste se obtiene un valor alto tanto para la TPR como FPR, en cambio para el valor máximo ambas mediciones disminuyen. Solo para el parámetro de ajuste de umbral para la señal de tipo Inductancia interna del plasma se logró encontrar un punto donde se mejora la performance del sistema, alcanzando una TPR del 96.25(%) y FPR de 5.8(%)

para los demás parámetros el entrenamiento acertó con el valor óptimo. Adicionalmente se destaca en las gráficas, mediante un hexagrama, el resultado obtenido tras el proceso de entrenamiento.

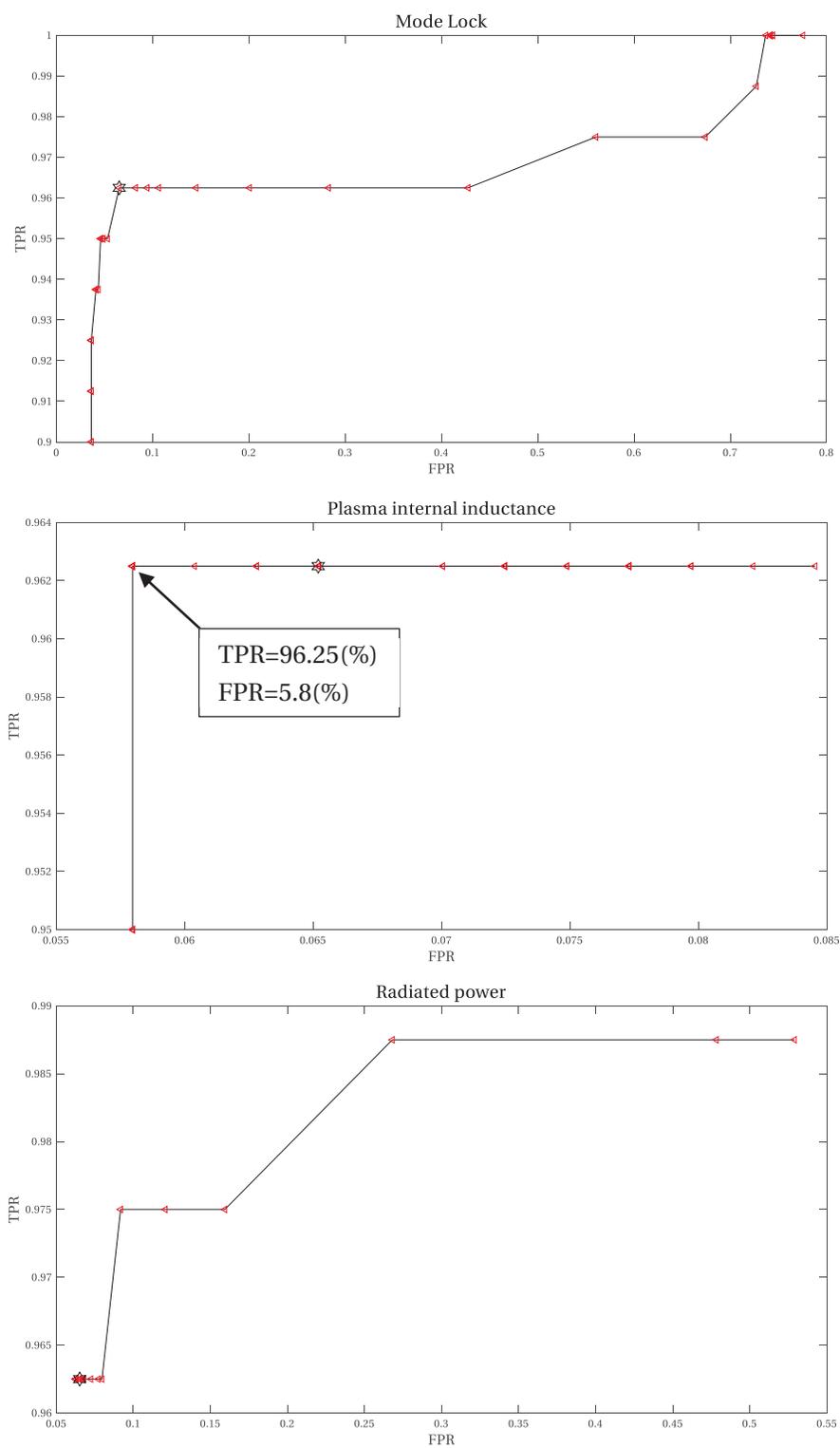


Figura 6-6.a: Curvas ROC para el criterio 1

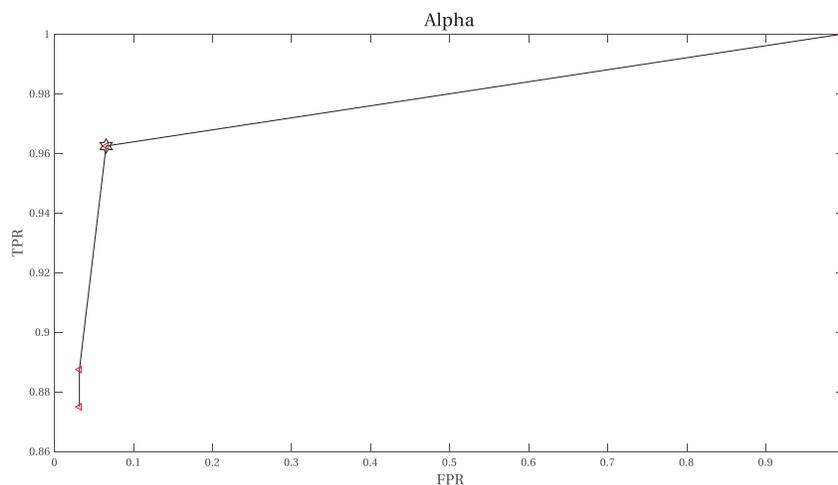


Figura 6-6.b: Curvas ROC para el criterio 1

Curva ROC – Criterio 2

Para el segundo criterio del predictor ninguna la variación en los parámetros produjo una mejora en los resultados, por lo que se puede aseverar que el proceso de optimización logró el mejor caso.

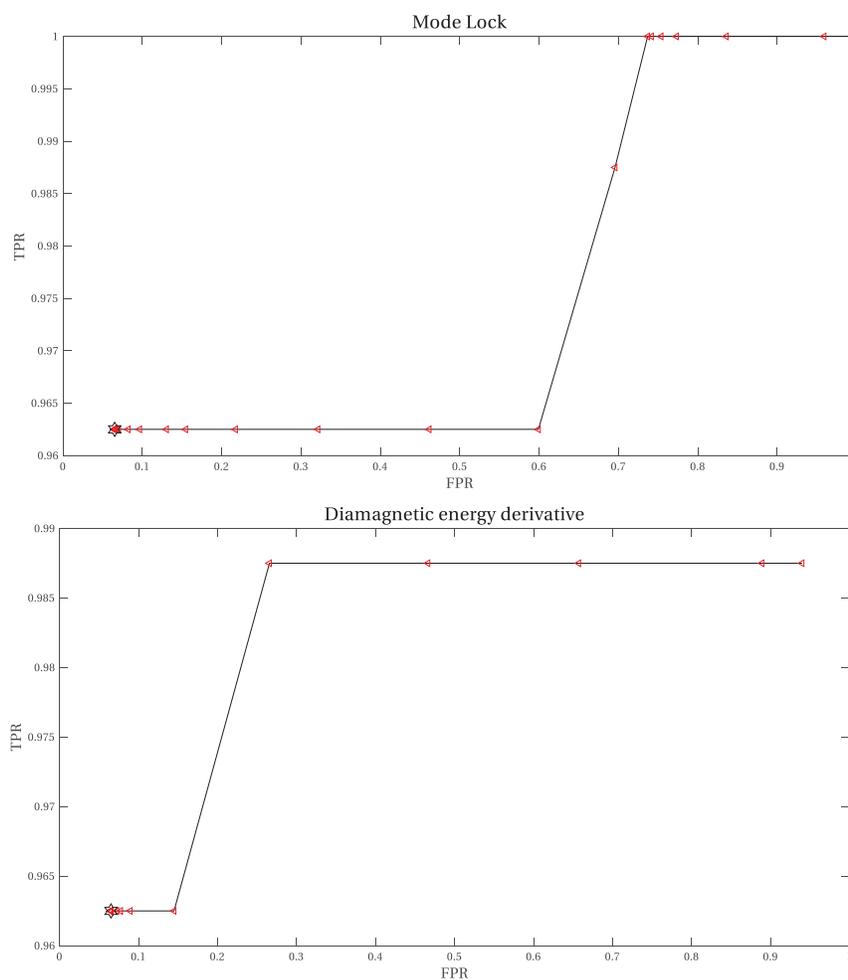


Figura 6-7.a: Curvas ROC para el criterio 2

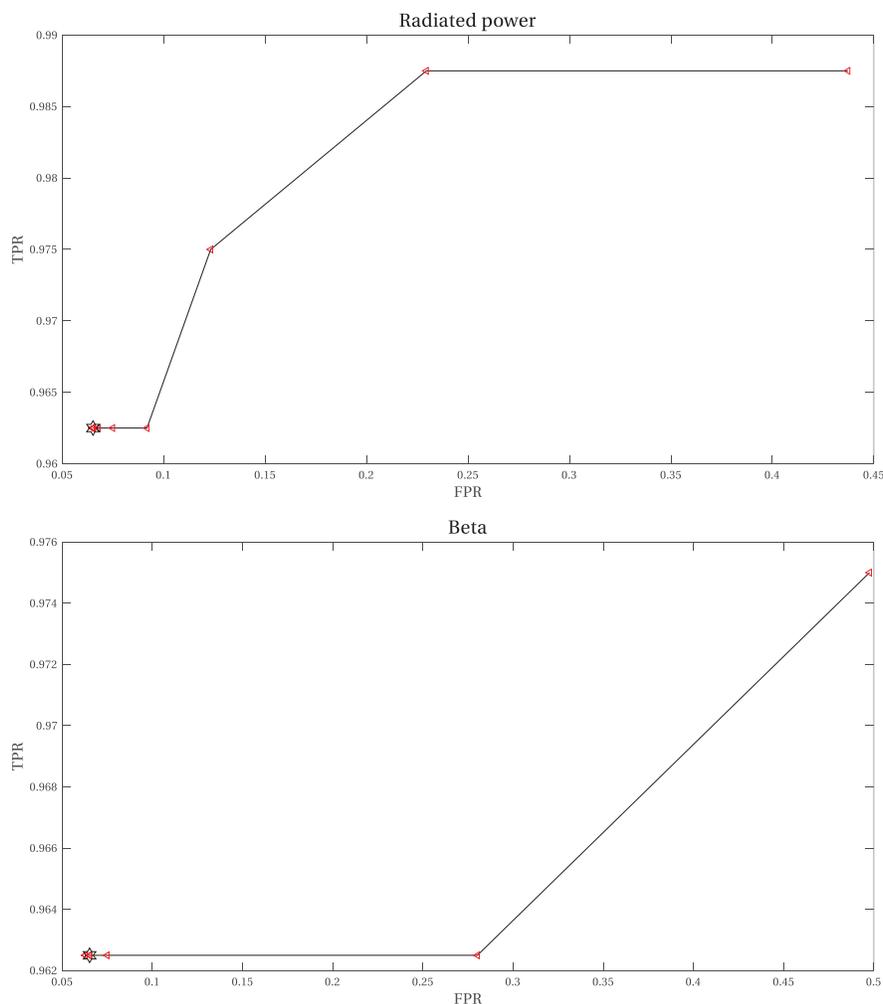


Figura 6-7.b: Curvas ROC para el criterio 2

Curva ROC – Criterio 3

Para este último criterio se obtuvo un resultado similar al primero, donde una de las señales, en este caso para la de tipo Mode Lock, se encontró un valor con el cual se alcanza una TPR y FPR del 96.25(%) y 5.8(%) respectivamente. Respecto al resto de parámetro, las curvas ROC no mostraron que existiese un valor que mejore los resultados obtenidos en el entrenamiento.

Las variaciones en todos los parámetros generaron el mismo efecto en las tasas de medición, si uno de los parámetros disminuye la TPR y FPR podrían solamente aumentar y si el parámetro aumenta la TPR y FPR necesariamente se mantienen o decrecen. Lo anterior se puede expresar matemáticamente de acuerdo a la ecuación (6-1).

$$\Delta^{\pm}X \Rightarrow \Delta^{\mp}TPR \wedge \Delta^{\mp}FPR \quad (6-1)$$

Donde X corresponde a cualquier parámetro del sistema DPAD.

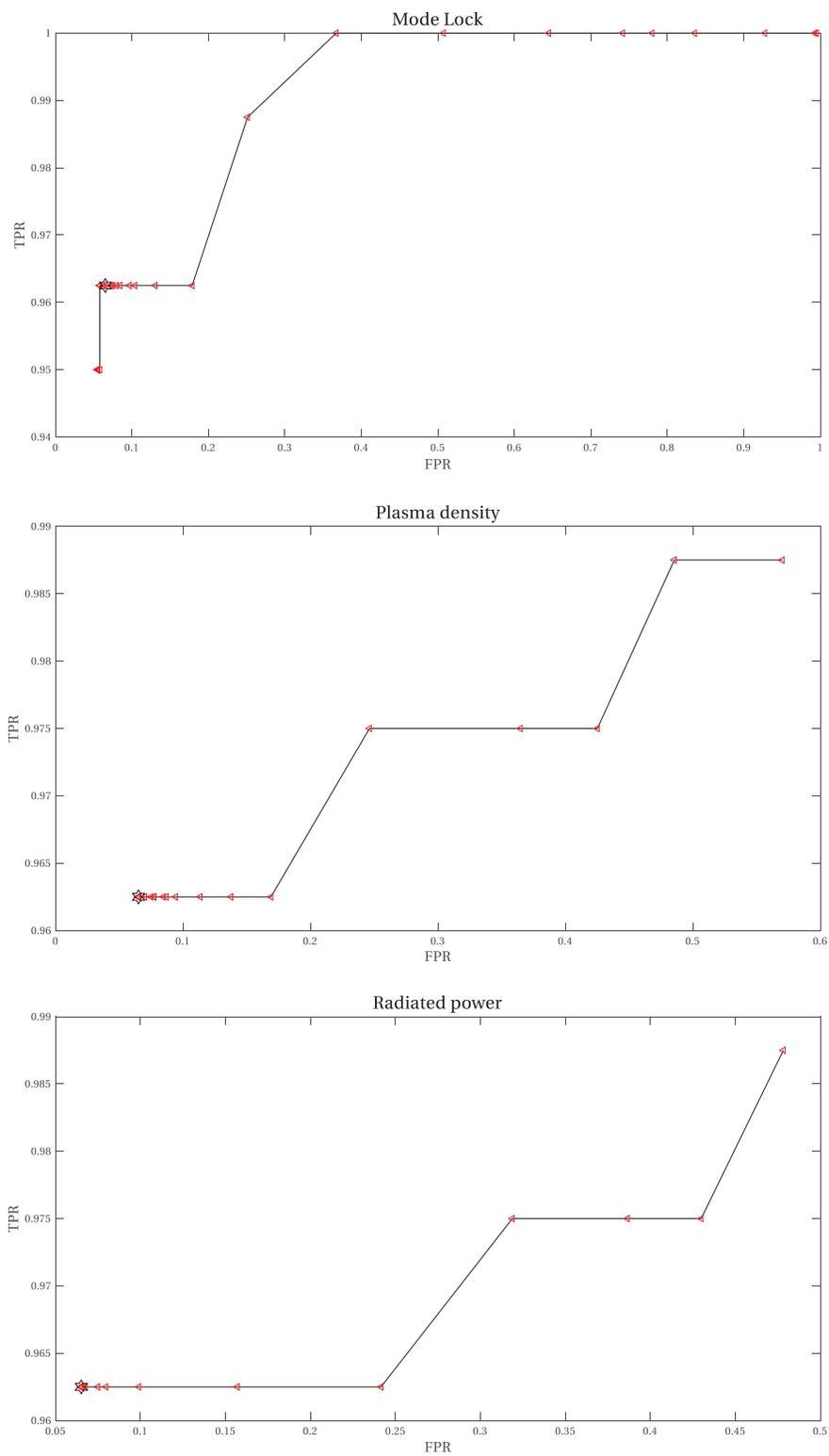


Figura 6-8.a: Curvas ROC para el criterio 3

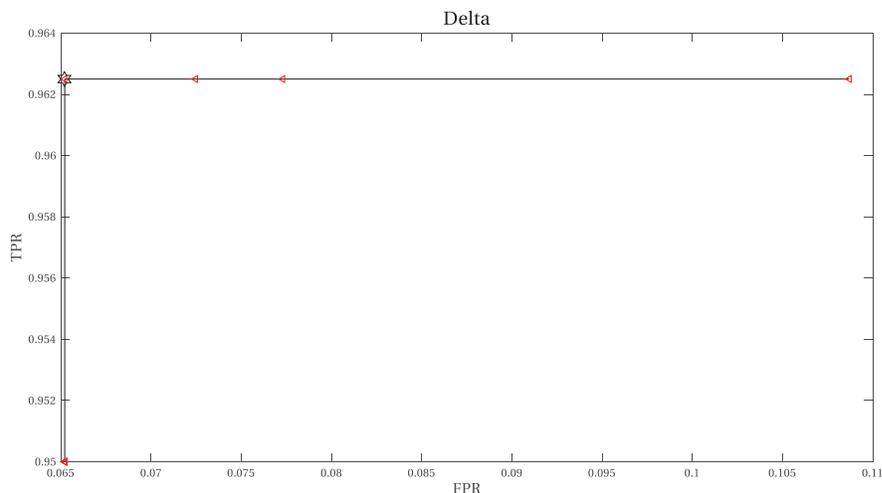


Figura 6-8.b: Curvas ROC para el criterio 3

Finalmente los resultados obtenidos en el mejor de los casos fueron de 96.25(%) para la TPR y de 5.8(%) la FPR, es decir, hubo una disminución de 0.72(%) en la tasas de falsos positivos sin afectar la tasa de verdaderos positivos. Adicionalmente, es destacable el hecho que el sistema se comportó de acuerdo a lo que se esperaba previamente, esto es gracias a la sencillez del sistema (caja transparente) y que cada uno de los parámetros tienen un sustento lógico. Como conclusión de este capítulo es posible señalar que el sistema propuesto es capaz de detectar precursores disruptivos, logrando tasas de acierto comparable e incluso mejores que sistema predictor altamente complejos. Por añadidura, es importante resaltar que este sistema basado en detección de anomalías puede ser implementado para detectar cualquier fenómeno que suceda en un experimento de fusión, esto posee mayor relevancia si se toma en cuenta que para ITER se espera que los disparos duren en torno a los 30(*min*) y se muestreen cerca de 6 millones de señales, por lo que discriminar lapsos de tiempo donde suceden fenómenos anómalos es de amplio interés científico.

Discusión y conclusiones

Se estima que la demanda energética crezca en 30(%) para el año 2040, los combustibles fósiles por su parte se están agotando a un ritmo vertiginoso y sumado a esto, la contaminación medioambiental debido a los combustibles fósiles (principales métodos de generación de energía) está en un estado crítico, estas tres observaciones llevan a concluir que es de vital importancia nuevas formas de generación de energía para realizar un cambio de paradigma respecto al escenario energético mundial.

De acuerdo a lo anterior surge la Fusión Nuclear como un método de generación de energía debido a que es capaz de manejar elevados niveles de potencia, los combustibles utilizados Deuterio y Tritio son en la práctica ilimitados y además no genera ningún tipo de residuo nocivo para el ambiente. A pesar de lo atractivo que suena, la alta inestabilidad que se genera en el confinamiento del plasma provoca disrupciones, lo cual a su vez conlleva que no sea viable económicamente su implementación y además que aún las energías obtenidas al final del proceso son menores a las que se utilizan para alcanzar los altísimos niveles de temperatura en el plasma.

La mitigación o evitación de las disrupciones es trascendental para la implementación efectiva de los dispositivos de fusión nuclear, existe una serie de técnicas relacionadas con estos tópicos, pero requieren de tiempos de anticipación, es decir que es necesario predecir si el plasma tendrá una inestabilidad súbita o no y además la predicción debe venir acompañada de suficiente tiempo para realizar alguna de las operaciones.

Por consiguiente, se elaboró un sistema capaz de predecir disrupciones basado en detección de anomalías, el cual fue construido a través de dos etapas. En una primera etapa se realiza la detección de outlier a través de un umbral definido por cada tipo de señal y el lanzamiento de una alarma disruptiva dependía de dos criterios: el criterio 1 gatilla la alarma cuando una cantidad establecida de muestras anómalas se generan en diferentes tipos de señales para un mismo instante y el criterio 2 se activa cuando se detectan de forma consecutiva una cantidad de outlier para un mismo tipo de señal. En la segunda etapa del predictor se amplió la cantidad de umbrales de detección, de manera que no fueran solo por tipo de señal, sino que también por criterio, de manera que estos se adaptarán de manera más fina. Adicionalmente se incluyó un nuevo criterio, el cual dispara una alarma cuando la suma de todas las anomalías detectadas, independiente del tipo de señal y dentro de una ventana temporal superan un umbral.

En definitiva las modificaciones lograron una mejora de la tasa de acierto desde 90(%) a 96.25(%), pero un deterioro de las falsas alarmas desde 1.77(%) hasta un 6.52(%), en definitiva las modificaciones lograron que el predictor se acercara de mejor manera a los requerimientos establecidos por ITER y que los resultados fueran comparables e incluso mejores que sistemas predictores más complejos. De esta manera se destaca como conclusión la sencillez del sistema DPAD, siendo una cualidad importante ya que no se comporta como caja negra, lo cual deriva en la posibilidad de modificar parámetros del sistema y anticiparse a las posibles variaciones en las tasas, es decir que se puede modificar la sensibilidad del predictor sin un reentrenamiento. Lo último se vio reflejo en las curvas ROC construidas para el análisis. Continuando con el estudio de los resultados, el predictor propuesto entregó información respecto a qué tipo de señales tienen mayor grado de participación en la generación de anomalías gatilladoras de alarma, donde sobresalen las señales Mode Lock, Plasma density y Plasma internal inductance.

Adicionalmente el predictor puede ser enfocado a la búsqueda de anomalías que tengan relación con otros fenómenos que suceden en el plasma, es decir que puede ser generalizado y no enfocarse estrictamente la predicción de disrupciones, este punto toma una alta relevancia si se considera que en dispositivos de fusión como JET no se analizan ni procesan más del 10(%) de los datos y además, se espera que en ITER la cantidad de datos que se generen anualmente sea entorno a los 100(PB).

Uno de los requerimientos que tienen la nueva generación de dispositivos tokamak como ITER y DEMO, es el hecho de no poseer datos al inicio de sus operaciones, para simular dicha situación se elaboró el método de entrenamiento “from scratch” o entrenamiento desde cero (que se adjunta en el apéndice de este trabajo), en donde secuencialmente se van realizando diferentes modelos predictores a medida que el dispositivo va generando nuevos datos. Los resultados obtenidos a partir de este método en el predictor propuesto, muestran una tasa de acierto máxima de 89(%) y mínima de falsas alarmas de 7(%) aproximadamente, respecto al contraste con modelos como APODIS y Veen Predictor se observó que aunque el comportamiento es similar, es necesario realizar mejoras al sistema predictor. De acuerdo a los puntos comentados anteriormente se definen una serie de futuros trabajos a realizar.

Predictor del tiempo a la disrupción

En el momento en que se gatilla una alarma disruptiva no se tiene conocimiento del tiempo requerido para que esto suceda, de manera que no se sabría si aplicar métodos de mitigación o evitación, los cuales requieren WT muy diferentes. Esta problemática no ha sido abordada de manera muy exhaustiva debido a su alta complejidad, por lo que encontrar un modelo que indique el tiempo a la disrupción o determinar precursores que se relacionen con esto sería una investigación de alto interés científico. De acuerdo a lo anterior se plantea como trabajo futuro construir un modelo que también sea capaz de predecir el tiempo al fenómeno súbito. Es así como se plantea ampliar el sistema predictor previamente desarrollado mediante la implementación de un bloque denominado PTD (Predictor of Time to Disruption). Una propuesta es utilizar ventanas previas al momento en que se gatilla la alarma para construir un clasificador basado en algoritmos de Deep Learning.

Tipos de interrupciones

Es posible encontrar trabajos donde realizan una clasificación de las diferentes causas o tipos de interrupciones, entonces si fuese posible dividir la data de acuerdo a estos criterios y realizar clasificadores más específicos y, por ende, con un ajuste más fino, sería viable generar predictores más confiables y con mejores tasas de acierto. Adicionalmente, mirado desde otra perspectiva, anticiparse al posible tipo de interrupción permitiría aplicar técnicas específicas de mitigación o evitación.

Incertidumbre en la predicción

A la hora de realizar la predicción es importante que venga acompañada con una medición según qué tan segura sea, es decir que en nuestro caso a la hora de predecir una interrupción se debe determinar cuan confiable y creíble es la clasificación, de manera que se pueda ir más allá de simples clasificaciones puntuales, sino que tome en cuenta la incertidumbre del proceso. En palabras simples la credibilidad hace referencia a que tan segura es la etiqueta predicha y la confianza que tan distinta es la muestra predicha con las demás etiquetas. Un método de posible a implementar y que se puede abordar mediante el enfoque que se plantea en este trabajo es el de Predictores Conformales [45].

Nuevo método de entrenamiento

El entrenamiento a través de la implementación de algoritmos genéticos no entrega el óptimo general del sistema y adicionalmente el proceso puede converger en diferentes soluciones debido a la serie de decisiones aleatorias que se realizan, según lo anterior modificar el entrenamiento permitiría atacar estos puntos e incluso disminuir los tiempos de entrenamiento. Según esto se plantea como trabajo futuro la implementación de métodos de aprendizaje por reforzamiento.

Bibliografía

- [1] A. Eddington, *The internal constitution of the stars*, Cambridge: Cambridge University Press, 1988.
- [2] M. Lehnen, A. Alonso, G. Arnoux, N. Baumgarten, S. Bozhenkov, S. Brezinsek y a. others, «Disruption mitigation by massive gas injection in JET,» *Nuclear fusion*, vol. 51, nº 12, p. 123010, 2011.
- [3] Organización de las Naciones Unidas, Department of Economic and Social Affairs, «World Population Prospect: The 2017 Revision,» New York, 2017.
- [4] Agencia Internacional de la Energía, «World Energy Outlook 2018,» Washington, 2018.
- [5] Consejo Mundial de la Energía, «World Energy Resources,» Londres, 2013.
- [6] Organización Mundial de la Salud, «Centro de Prensa OMS,» 2016 Junio 29. [En línea]. Available: <http://www.who.int/mediacentre/factsheets/fs266/es/>. [Último acceso: 14 Diciembre 2018].
- [7] J. Lawson, «Power from Nuclear Fusion,» *Nature*, vol. 180, pp. 780-782, 1957.
- [8] M. Nieto, *Procesos de ignición en fusión por confinamiento inercial*, Univ de Castilla La Mancha, 2011.
- [9] S. Glenzer, B. MacGowan, P. Michel, N. Meezan, S. LJ, D. SN y and others, «Symmetric inertial confinement fusion implosions at ultra-high laser energies,» *Science*, vol. 327, nº 5970, pp. 1228-1231, 2010.
- [10] J. L. Johnson, «Stellarator and Heliotron Devices,» *Nuclear Fusion*, vol. 39, nº 2, pp. 293-294, 1998.

-
- [11] J. H. Harris, «A flexible heliac configuration,» *Nuclear Fusion*, vol. 25, nº 5, pp. 623-629, 1985.
- [12] M. Kubic, «Review of plasma parameters of the JET tokamak in various regimes of its operation,» *Bachelor thesis, Czech Technical University, Prague*, pp. 14-19, 2007.
- [13] ITER, «Building ITER,» 2017.. [En línea]. Available: <https://www.iter.org>. [Último acceso: 14 Diciembre 2018].
- [14] S. Ciattaglia, G. Federici, L. Barucca, A. Lampasi, S. Minucci y I. Moscato, «The European DEMO fusion reactor: Design status and challenges from balance of plant point of view,» de *Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), 2017 IEEE International Conference on*, IEEE, 2017, pp. 1-6.
- [15] M. Kikuchi, K. Lackner y M. Q. Tran, *Fusion Physics*, Vienna: International Atomic Energy Agency (IAEA), 2012.
- [16] J. Wesson, «Hydromagnetic stability of tokamaks,» *Nuclear Fusion*, vol. 18, nº 1, p. 87, 1978.
- [17] J. Wesson, R. Gill, M. Hugon, F. Schüller, J. Snipes y a. others, «Disruptions in JET,» *Nuclear Fusion*, vol. 29, nº 4, p. 641, 1989.
- [18] A. Savtchikov, *Mitigation of disruptions in a tokamak by means of large gas injection*, Zentralbibliothek: Forschungszentrum Jülich in der Helmholtz-Gemeinschaft, 2004.
- [19] P. De Vries, M. Johnson, B. Alper, P. Buratti, T. Hender, H. Koslowski, V. Riccardo, J.-E. Contributors y a. others, «Survey of disruption causes at JET,» *Nuclear Fusion*, vol. 51, nº 5, p. 053018, 2011.
- [20] M. Bakhtiari, G. Olynyk, R. Granetz, D. Whyte, M. Reinke, K. Zhurovich y V. Izzo, «Using mixed gases for massive gas injection disruption mitigation on Alcator C-Mod,» *Nuclear Fusion*, vol. 51, nº 6, p. 063007, 2011.
- [21] G. Pautasso, K. Buchl, J. Fuchs, O. Gruber, A. Herrmann, K. Lackner, P. Lang, K. Mast, M. Ulrich y H. Zohm, «Use of impurity pellets to control energy dissipation during disruption,» *Nuclear fusion*, vol. 36, nº 10, p. 1291, 1996.
- [22] N. Commaux, L. R. Baylor, S. K. Combs y a. others, «Novel rapid shutdown strategies for runaway electron suppression in DIII-D,» *Nuclear Fusion*, vol. 51, nº 10, p. 103001, 2011.
- [23] B. Esposito, G. Granucci, P. Smeulders, S. Nowak, J. Martin-Solis, L. Gabellieri y a. others, «Disruption avoidance in the Frascati Tokamak Upgrade by means of

- magnetohydrodynamic mode stabilization using electron-cyclotron-resonance heating,» *Physical review letters*, vol. 100, n° 4, p. 045006, 2008.
- [24] J. Vega, S. Dormido-Canto, J. M. López, A. Murari, J. Ramírez, R. Moreno, M. Ruiz y D. Alves, «Results of the JET real-time disruption predictor in the ITER-like wall campaigns,» *Fusion Engineering and Design*, vol. 88, n° 6-8, pp. 1228-1231, 2013.
- [25] P. De Vries, M. Johnson, I. Segui y J. E. Contributors, «Statistical analysis of disruptions in JET,» *Nuclear Fusion*, vol. 49, n° 5, p. 055011, 2009.
- [26] G. Rattá, Vega, Jesús, A. Murari, G. Vagliasindi, M. Johnson, P. De Vries y J. E. Contributors, «An advanced disruption predictor for JET tested in a simulated real-time environment,» *Nuclear Fusion*, vol. 50, n° 2, p. 025005, 2010.
- [27] R. Yoshino, «Neural-net predictor for beta limit disruptions in JT-60U,» *Nuclear fusion*, vol. 45, n° 11, p. 1232, 2005.
- [28] C. Windsor, G. Pautasso, C. Tichmann, R. Buttery, T. Hender, J. E. Contributors y a. others, «A cross-tokamak neural network disruption predictor for the JET and ASDEX Upgrade tokamaks,» *Nuclear fusion*, vol. 45, n° 5, p. 337, 2005.
- [29] B. Cannas, A. Fanni, P. Sonato, M. K. Zedda, J.-E. contributors y a. others, «A prediction tool for real-time application in the disruption protection system at JET,» *Nuclear Fusion*, vol. 47, n° 11, p. 1559, 2007.
- [30] A. Murari, G. Vagliasindi, P. Arena, L. Fortuna, O. Barana, M. Johnson, J.-E. Contributors y a. others, «Prototype of an adaptive disruption predictor for JET based on fuzzy logic and regression trees,» *Nuclear Fusion*, vol. 48, n° 3, p. 035010, 2008.
- [31] J. Vega, A. Murari, S. Dormido-Canto, R. Moreno, A. Pereira, S. Esquembri y J. Contributors, «Real-time anomaly detection for disruption prediction: the JET case,» *Nuclear Fusion*, 2014.
- [32] J. M. López, J. Vega, D. Alves, S. Dormido-Canto, A. Murari, G. Arcas, R. Felton y J. Ramírez, «Implementation of the disruption predictor APODIS in JET's real-time network using the MARTe framework,» *IEEE Transactions on Nuclear Science*, vol. 61, n° 2, pp. 741-744, 2014.
- [33] B. Cannas, R. Delogu, A. Fanni, P. Sonato, M. K. Zedda, J.-E. contributors y a. others, «Support vector machines for disruption prediction and novelty detection at JET,» *Fusion engineering and design*, vol. 82, n° 5-14, pp. 1124-1130, 2007.
- [34] G. Pautasso, C. Tichmann, S. Egorov, T. Zehetbauer, O. Gruber, M. Maraschek, K.-F. Mast, V. Mertens, I. Perchermeier, G. Raupp y a. others, «On-line prediction and mitigation of disruptions in ASDEX Upgrade,» *Nuclear Fusion*, vol. 42, n° 1, p. 100, 2002.

-
- [35] A. Murari, M. Lungaroni, E. Peluso, P. Gaudio, J. Vega, S. Dormido-Canto, M. Baruzzo, M. Gelfusa y J. Contributors, «Adaptive predictors based on probabilistic SVM for real time disruption mitigation on JET,» *Nuclear Fusion*, vol. 58, nº 5, p. 056002, 2018.
- [36] J. Vega, A. Murari, S. Dormido-Canto, R. Moreno, A. Pereira, A. Acero, J.-E. Contributors y a. others, «Adaptive high learning rate probabilistic disruption predictors from scratch for the next generation of tokamaks,» *Nuclear Fusion*, vol. 54, nº 12, p. 123001, 2014.
- [37] M. Lungaroni, A. Murari, E. Peluso, J. Vega, G. Farias, M. Gelfusa y J. Contributors, «On the potential of ruled-based machine learning for disruption prediction on JET,» *Fusion Engineering and Design*, vol. 130, pp. 62-68, 2018.
- [38] D. M. Hawkins, *Identification of outliers*, Dordrecht: Springer, 1980.
- [39] V. Chandola, A. Banerjee y V. Kumar, «Anomaly detection: A survey,» *ACM computing surveys (CSUR)*, vol. 41, nº 3, p. 15, 2009.
- [40] W. S. Ruiz, «Técnicas de Minería de Datos Aplicadas en la Detección de Fraude: Estado del Arte,» *Universidad Nacional de Colombia*, 2006.
- [41] S. Mallat, *A wavelet tour of signal processing: the sparse way*, Burlington: Academic press, 2008.
- [42] R. De Maesschalck, D. Jouan-Rimbaud y D. L. Massart, «The Mahalanobis distance,» *Chemometrics and intelligent laboratory systems*, vol. 50, nº 1, pp. 1-18, 2000.
- [43] E.-G. Talbi, *Metaheuristics: from design to implementation*, New Jersey: John Wiley & Sons, 2009.
- [44] The MathWorks, Inc., «Genetic Algorithm Options,» [En línea]. Available: <https://la.mathworks.com/help/gads/genetic-algorithm-options.html#f7820>. [Último acceso: 14 Diciembre 2018].
- [45] V. Vovk, A. Gammerman y G. Shafer, *Conformal prediction*, Boston: Springer, 2005.
- [46] R. Moreno Salinas, *Advanced techniques of disruption prediction, application to JET and extrapolation to ITER*, Madrid: Universidad Nacional de Educación a Distancia (España). Escuela Técnica Superior de Ingeniería Informática, 2015.

A Entrenamiento desde cero

A.1 Predictores desde cero

La nueva generación de dispositivos tokamak como ITER y DEMO es fundamental implementar sistemas que predigan eficientemente los fenómenos disruptivos y se debe considerar el hecho de que no existe información previa para construir modelos predictores. De esta manera los sistemas DPFS (Disruptive Predictor From Scratch) [36], predictores desde cero en español, son de vital importancia para el futuro de la Fusión Nuclear.

En el método DPFS, el primer modelo predictor se entrena una vez generado el primer disparo disruptivo, generalmente en los dispositivos tokamak en el inicio de sus operaciones, se producen varias descargas safe antes de la primera disruptiva (k_1 número de descargas antes de la primera disruptiva), las $k_1 + 1$ descargas (k_1 safe y 1 disruptiva) conforman el primer conjunto de entrenamiento para el sistema predictor. Una vez encontrado el valor de los parámetros del sistema que optimizan su aptitud se testea su comportamiento con la siguiente descarga $k_1 + 2$. El último conjunto de entrenamiento y los parámetros obtenidos siguen siendo válidos hasta que se falle una alarma, en dicho caso es necesario volver a entrenar el predictor a través de un nuevo conjunto de entrenamiento, en el cual se incluyen todas las descargas tanto safe como disruptivas previas y en la que se falló. En la Tabla A-1 se muestra mediante un pseudocódigo la implementación del método DPFS con el predictor propuesto.

A.1.1 Implementación método DPFS

Para implementar el algoritmo que se indica en la Tabla A-1, se realizó ordenamiento de los disparos de forma cronológica. Respecto al proceso de entrenamiento, que nuestro caso es la búsqueda de los parámetros de acuerdo a la optimización mediante algoritmos genéticos, se implementó el sistema DPAD en su primera etapa y se tuvo que reducir o discretizar el espacio de búsqueda por el excesivo tiempo que le tomaba al sistema en converger, entonces se estableció que los parámetros d_i no pueden variar en menos de 0.01, sumado a lo anterior se estableció un límite de la función objetivo igual a 0.9025 y cuando no se logra una mejora en la optimización en a lo más 3 hora. En definitiva, el problema de optimización es igual a la que fue descrito en el planteamiento matemático (5-2).

Tabla A-1 Algoritmo de entrenamiento desde cero

Método DPFS
Inicio de la operación del dispositivo Tokamak
Para k_1 descargas no-disruptivas (safe)
Primer descarga disruptiva ($k_1 + 1$)
Primer conjunto de entrenamiento
Entrenamiento, determinación de parámetros (d_i, α y β)
Bucle de predicción
Inicio de la descarga
Predicción = salida sistema predictor
if (<i>Predicción == disruptivo & Descarga es disruptiva</i>)
Predicción correcta
TP=TP+1
break
else if (<i>Predicción == disruptivo & Descarga es safe</i>)
Predicción fallida
FP=FP+1
break
end if
Fin de descarga
if (<i>Descarga es disruptiva</i>)
Alarma perdida
Nuevo conjunto de entrenamiento
Entrenamiento, determinación de parámetros (d_i, α y β)
end if
Fin Bucle de predicción
Evaluación TPR y FPR

Este enfoque para el entrenamiento ha sido implementado en diferentes predictores, en los cuales destaca APODIS (sección 6.3) y Venn Predictor, este último es descrito brevemente en el siguiente apartado.

A.1.2 Veen Predictor

Los modelos predictores de interrupciones orientados a datos están basados en técnicas de clasificación automática como SVM (APODIS), redes neuronales, SOM (Mapa auto-organizado), k vecinos más próximos, entre otros. Estos modelos se generan a partir de 2 pasos fundamentales, inducción y deducción.

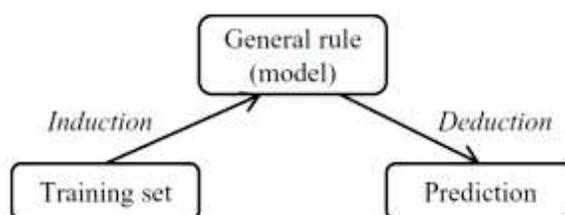


Figura A-1: Pasos de un sistema clasificador basado en datos [36]

De acuerdo a lo anterior es plantea la interrogante, ¿Es posible predecir sin la necesidad de utilizar una regla general? Los predictores Venn son predictores probabilísticos que proporcionan una barra de error de probabilidad para cada predicción individual. Estos predictores pertenecen a la familia de predictores conformes, los cuales, a diferencia de otros métodos de aprendizaje automático, brindan información sobre su propia fiabilidad y confiabilidad, además no siguen los pasos inductivo-deductivo (Figura A-1), sino que las predicciones conformes usan un marco alternativo que hace predicciones de forma secuencial, basando cada nueva predicción en todas

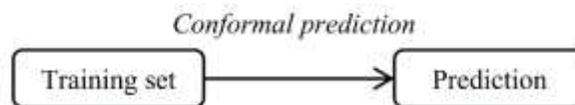


Figura A-2: Los predictores conformales no siguen los pasos inducción y deducción [36]

las muestras anteriores (Figura A-2).

Los predictores de Venn son clasificadores probabilísticos que determinan la probabilidad de cada predicción individual. Esto implica que no hacen predicciones simples, pero proporcionan un intervalo de probabilidad sobre cuán precisa y confiable es cada predicción. En términos de predictores de interrupción, esto significa que cada predicción disruptiva / no disruptiva se califica con un intervalo de probabilidad que se usa como el nivel de confianza de la predicción.

En palabras simples el sistema funciona inicialmente con ventanas temporales a las cuales se les extrae una cantidad de características, una vez producido la primera descarga disruptiva el método es capaz de emplearse. Cabe destacar que se realiza un balance entre las muestras, es decir, para la primera predicción se utiliza como ventana con carácter disruptivo, aquel vector de características que se obtuvo anterior a la interrupción, para las ventanas de clase no-disruptiva se calcula los valores medios de las diferentes características entre la primera descarga no disruptiva y la última antes de la primera interrupción. A medida que se desarrolla el método DPFS, se van agregando nuevas ventanas de clases disruptiva y safe.

A.2 Resultados entrenamiento desde cero sistema DPAD

La base de datos de las descargas fue ordenada cronológicamente y se construyó el primer predictor cuando sucedió la primera interrupción, siguiendo el algoritmo descrito anteriormente. En la Figura A-3 se muestra el desarrolló cronológico de las tasas de verdaderos positivos TPR (disparos disruptivos bien predichos), falsos positivos FPR (disparo safe predicho como disruptivo) y se destacan en que disparo se realizó el entrenamiento del predictor.

Se aprecia en los resultados claramente como a medida que se agrega información de nuevos disparos, ambas tasas se estabilizan. Desde el disparo disruptivo número 60 la TPR varía en torno 80(%) y 90(%), alcanzado un máximo de aproximadamente 89(%)

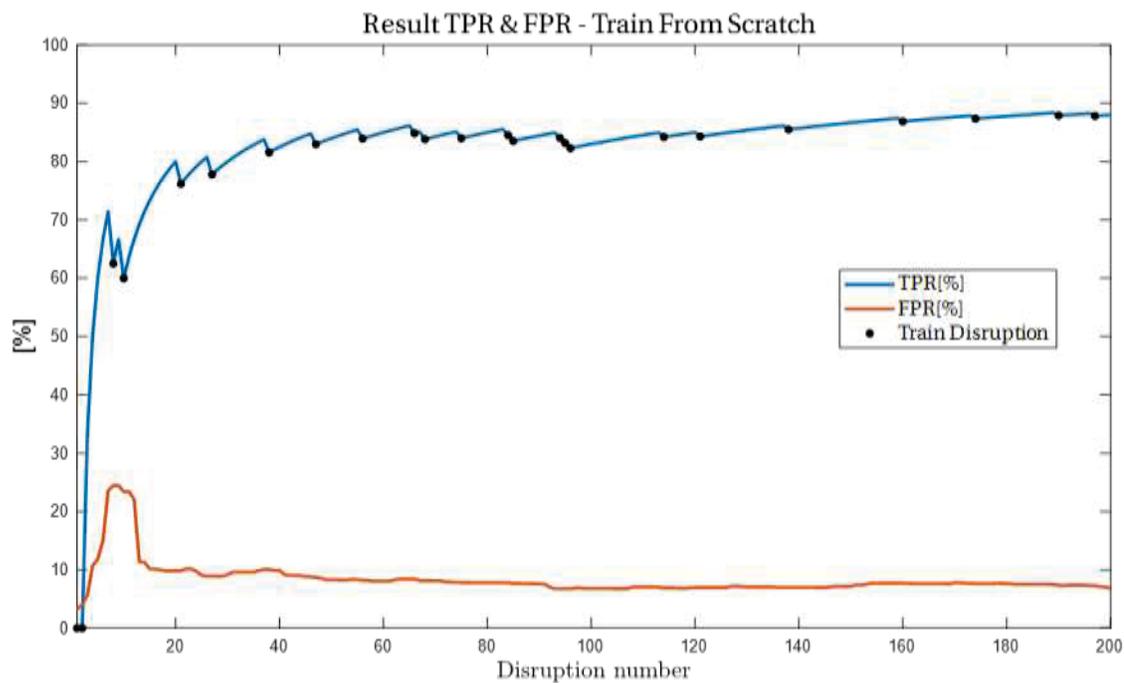


Figura A-3: Desarrollo TPR y FPR en entrenamiento desde cero

A través del proceso de entrenamiento “from scratch” se realizaron 22 procesos de optimización, de los cuales se obtuvieron los resultados que se detallan en la Tabla A-2.

Tabla A-2 Resultados TPR y FPR de cada entrenamiento

# Disrupción	TPR (%)	FPR(%)	# Disrupción	TPR(%)	FPR (%)
1	0	3.2	84	84.5	7.8
2	0	4	85	83.5	7.7
8	62.5	24.4	94	84	6.7
10	60	23.4	95	83.2	6.7
21	76.2	9.8	96	82.3	6.7
27	77.8	8.9	114	84.2	6.9
38	81.6	10.1	121	84.3	6.9
47	83	8.7	138	85.5	7
56	83.9	8.2	160	86.9	7.7
66	84.9	8.4	174	87.4	7.6
68	83.8	8.1	190	87.9	7.3
75	84	7.8	197	87.8	7.3

A.3 Comparativa

APODIS y Veen Predictor [36] son sistemas que han sido entrenados a través del método DPFS y utilizaron la misma base de datos con que se desarrolla este trabajo, por lo que realizar una comparación es provechoso para concluir respecto al desarrollo del sistema.

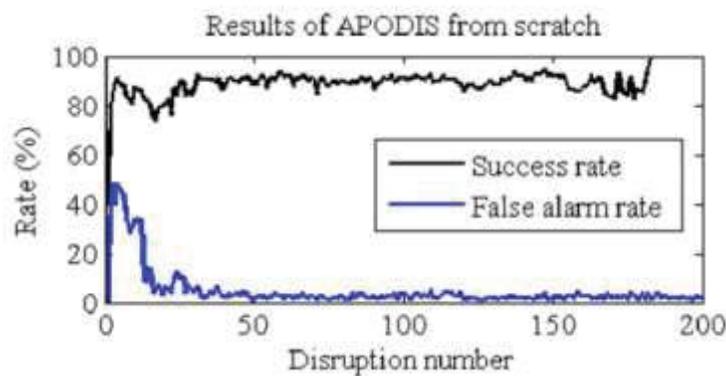


Figura A-5 Resultados de APODIS desde cero [36]

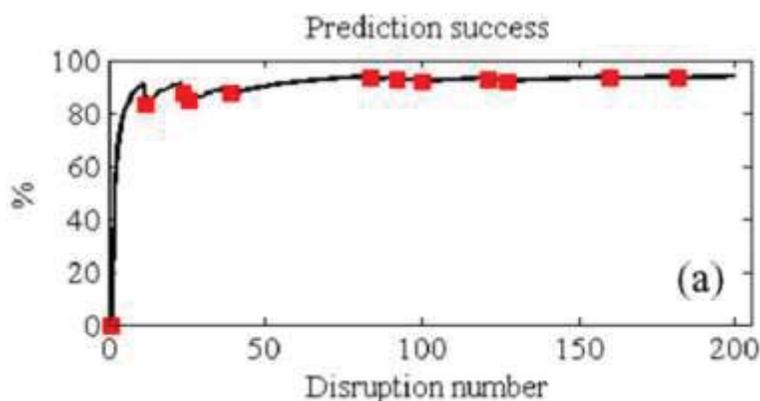


Figura A-4: Resultados de Venn Predictor desde cero [36]

Las figuras anteriores muestran el desarrollo del proceso de entrenamiento desde cero de ambos sistemas, alcanzado tasas de acierto comparables (APODIS 93.5(%) y Venn Predictor 94(%)) y una estabilización de las curvas en el disparo disruptivo número 40.

La comparación indica que, aunque existe una diferencia en torno al 5(%) en la tasa de acierto y que la estabilización de la TPR se genere luego de 60 disparos disruptivos aproximadamente (20 disruptions después de que se estabilizan los modelos citados), los resultados del sistema propuestos son aceptables, ya que como se dijo previamente el predictor basado en detección de anomalías tiene un alto grado de simpleza y adicionalmente no se evaluó el sistema completo debido al alto gasto temporal involucrado.