



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Biodiversidad y Biología Experimental

“Aproximación computacional a la caracterización del comportamiento de alimentación de un insecto vector”

Tesis presentada para optar al Título de Doctor/a de la Universidad de Buenos Aires en el área de Ciencias Biológicas

Lic. Héctor Luis Salas Morales

Director de tesis: Romina B. Barrozo

Co-director de tesis: Pablo E. Schilman

Consejero de Estudios: Gabriel Manrique

Lugar de trabajo: Laboratorio de Neuroetología de Insectos y Laboratorio de Fisiología de Insectos. DBBE, FCEN, IBBEA, UBA-CONICET.

Fecha de defensa: Buenos Aires, 11 de diciembre del 2023

Resumen

La vinchuca *Rhodnius prolixus* (Heteroptera: Reduviidae: Triatominae) es un insecto hematófago, vector de la enfermedad de Chagas en América Latina y modelo en fisiología de insectos. Su comportamiento alimenticio consta de dos fases importantes: 1- prueba del alimento, y 2- ingestión. En la fase de prueba, el insecto ingiere una porción del alimento y decide si continúa o no con la siguiente fase, donde ocurre la verdadera ingestión. Ambas fases están moduladas por la composición química del alimento que las vinchucas son capaces de identificar. En esta tesis, analizamos el comportamiento alimenticio de estos insectos al proporcionarles diferentes compuestos en la solución alimenticia. Dicho análisis fue posible gracias al desarrollo de un programa computacional de análisis del comportamiento de alimentación, que denominamos MyoRec.

Primeramente, analizamos el comportamiento alimenticio frente a compuestos de valencias opuestas, *i.e.* apetitivo *versus* aversivo, definiendo y estableciendo las variables comportamentales que permiten caracterizar la ingesta. Esto fue posible gracias al registro de la actividad eléctrica de los músculos succionadores que se encuentran en la cabeza de los insectos, responsables de la ingestión. Este tipo de registros se denominan electromiogramas (EMGs).

Seguidamente, con el fin de automatizar el proceso de reconocimiento de patrones de alimentación, comparamos la eficacia de diferentes algoritmos de *machine learning* y *deep learning*. El algoritmo que arrojó los mejores resultados predictivos fue la Red Neuronal Convolutiva (CNN, del inglés *Convolutional Neural Network*), con una exactitud de 0,974. Posteriormente, se desarrollaron dos aplicaciones que permiten el cálculo de las variables relevantes en la caracterización del comportamiento de alimentación: el Detector de Picos y el Calculador de Datos. Además, se desarrolló la interfaz gráfica que permite el análisis de los registros facilitando las tareas de organización, almacenamiento, visualización, procesamiento, análisis y comparación de los datos de cada EMG.

Finalmente, se utilizó MyoRec para analizar el comportamiento de *R. prolixus* frente a distintos compuestos apetitivos (glucosa, ATP, leucina, fenilalanina) y aversivos (cafeína, extracto de ajo, metil salicilato, quinina). Encontramos no solo diferencias significativas en el desempeño alimenticio frente a los diferentes compuestos, sino que MyoRec permitió el análisis comparado de manera rápida y sencilla de múltiples registros de alimentación.

Palabras clave: MyoRec, comportamiento de alimentación, estímulos apetitivos y aversivos, electromiograma, *machine learning*, *deep learning*, reconocimiento de patrones.

Abstract

Computational approach to the characterization of the feeding behavior of an insect vector

The kissing bug *Rhodnius prolixus* (Heteroptera: Reduviidae: Triatominae) is a hematophagous insect, a vector of Chagas disease in Latin America, and a model in insect physiology. Its feeding behavior consists of two important phases: 1- probing, and 2- ingestion. During the probing phase, the insect ingests a portion of the food and decides whether or not to proceed to the next phase, where the true ingestion occurs. Both phases are modulated by the chemical composition of the food that the kissing bugs can identify. In this thesis, we analyzed the feeding behavior of these insects by providing them with different compounds in the feeding solution. The analysis was possible thanks to the development of a computer program for feeding behavior analysis, which we call MyoRec.

First, we analyzed the feeding behavior in response to compounds with opposite valences, i.e. appetitive versus aversive, defining and establishing the behavioral variables that allow for the characterization of ingestion. This was achieved by recording the electrical activity of the sucking muscles located in the head of insects, responsible for ingestion. These recordings are known as electromyograms (EMGs)

Next, to automate the recognition pattern of the feeding process, we compared the effectiveness of different algorithms based on machine learning and deep learning models. The algorithm that yielded the best predictive results was the Convolutional Neural Network (CNN), with an accuracy of 0.974. Subsequently, two applications were developed that enabled the analysis of relevant variables in the characterization of feeding behavior: the Peak Detector and the Data Calculator. Additionally, a graphical interface was developed to facilitate the organization, storage, visualization, processing, analysis, and comparison of EMG data.

Finally, MyoRec was used to analyze the behavior of *R. prolixus* in response to different appetitive compounds (glucose, ATP, leucine, phenylalanine) and aversive compounds (caffeine, garlic extract, methyl salicylate, quinine). Not only did we find significant differences in feeding performance with respect to different compounds, but MyoRec also allowed for quick and easy comparative analysis of multiple feeding records.

Keywords: MyoRec, feeding behavior, appetitive and aversive stimuli, electromyogram, machine learning, deep learning, pattern recognition.

Agradecimientos

En el transcurso de este emocionante viaje hacia la culminación de mi tesis doctoral, me gustaría expresar mi profundo agradecimiento a las personas e instituciones que han sido fundamentales en este proceso.

En primer lugar, quiero agradecer a mi directora, Romina, por su guía constante, su paciencia inagotable y su sabiduría invaluable. Su compromiso con mi desarrollo académico ha sido inspirador y fundamental para alcanzar este logro.

A mi codirector, Pablo, le debo un agradecimiento especial por su valioso apoyo y orientación experta. Su perspicacia y conocimiento han enriquecido mi investigación.

Quiero expresar mi gratitud a todo el equipo del Laboratorio de Fisiología de Insectos, a su director Gabriel Manrique, y a Sebastián, Lizzy, y Tomas. Sus ideas, discusiones y colaboraciones han contribuido a mi experiencia y al desarrollo de esta tesis.

A mis padres, familiares y amigos, les agradezco por su apoyo inquebrantable a lo largo de este arduo proceso. En particular, quiero mencionar a mi madre, Syr, y a mi padre, Marco, por su amor y apoyo incondicional. A mi hermano, Javier, le agradezco por su aliento constante y amistad. A mi novia, Nathaly, le agradezco por su paciencia, comprensión y amor constante durante estos años.

No puedo pasar por alto la importancia de las instituciones que han respaldado mi investigación. A la Universidad de Buenos Aires, le agradezco por brindarme un entorno académico excepcional y recursos valiosos. Al CONICET, le agradezco por su apoyo financiero y la oportunidad de colaborar con destacados científicos en mi campo de estudio.

En resumen, esta tesis no habría sido posible sin el apoyo y la contribución de estas personas e instituciones. A todos ustedes, les estoy profundamente agradecido por ser parte de este viaje y por ayudarme a alcanzar esta meta académica.

Índice General

Resumen	2
Abstract	3
Agradecimientos	4
Índice General	5
Introducción general	9
I.1 Comportamiento de Búsqueda del Alimento	11
I.2 Comportamientos durante la Interacción o Contacto con el Hospedador	12
I.2.1 Etapa de Reconocimiento.....	12
I.2.2 Etapas de Ingestión.....	13
I.2.3 Fase de Muestreo	13
I.2.4 Fase de Verdadera Ingestión	14
I.3 Sistema Gustativo: Reconocimiento del Alimento	15
I.3.1 Sistema Gustativo o Quimiorrecepción de Contacto en Insectos	16
I.4 El Alimento de los Insectos Hematófagos: la Sangre	19
I.5 Registro del Comportamiento de Alimentación.....	20
I.5.1 Microscopía Intravital.....	20
I.5.2 Electropenetrografía.....	22
I.5.3 Electromiografía	22
I.6 Patrones de Onda en EPGs y EMGs	24
I.7 El Modelo de Estudio de esta Tesis	24
I.7.1 Aparato Bucal de <i>Rhodnius prolixus</i>	25
I.7.2 Estructura del Principal Músculo Suctor de <i>Rhodnius prolixus</i> : la Bomba Cibarial.....	26
I.8 Compuestos Presentes en el Alimento: Apetitivos versus Aversivos	28
I.9 Relevancia de esta Tesis	28
I.10 Objetivos de esta Tesis	29
I.10.1 Objetivo General.....	29
I.10.2 Objetivos Específicos	29
I.11 Hipótesis	30

I.11.1 Hipótesis 1	30
I.11.2 Hipótesis 2	30
I.11.3 Hipótesis 3	30
Capítulo 1: Caracterización del Comportamiento de Alimentación de <i>Rhodnius prolixus</i> a Compuestos Apetitivos y Aversivos	31
1.1 Introducción.....	31
1.2 Materiales y Métodos.....	32
1.2.1 Insectos.....	32
1.2.2 Soluciones.....	33
1.2.3 Alimentador Artificial	33
1.2.4 Sistema de Calentamiento.....	33
1.2.5 Sistema de Registro de la Actividad Muscular	33
1.2.6 Diseño Experimental.....	34
1.2.7 Análisis Estadístico.....	36
1.3 Resultados	36
1.3.1 Caracterización del Comportamiento de Alimentación	36
1.3.2 Comportamiento de Alimentación hacia una Solución Apetitiva y una Aversiva	37
1.4 Discusión.....	44
Capítulo 2. MyoRec: Programa Computacional para Optimizar el Estudio del Comportamiento de Alimentación	47
PARTE A: Identificación de Patrones de Comportamiento Mediante Modelos de Machine Learning y Deep Learning.....	47
2.A.1 Introducción.....	47
2.A.1.1 Antecedentes de Programas Computacionales Usados en el Análisis de Electromiogramas en Distintos Insectos	48
2.A.1.2 Modelos Matemáticos para el Reconocimiento de Patrones en el Identificador de Eventos de Verdadera Ingestión	49
2.A.2 Materiales y Métodos	50
2.A.2.1 Datos de Entrenamiento y Validación.....	50
2.A.2.2 Modelos de Clasificación Probados para el Identificador de Eventos de Verdadera Ingestión	52
2.A.2.3 Métricas.....	53

2.A.3 Resultados	54
2.A.3.1 Algoritmo de Refinación.....	57
2.A.4 Discusión	60
<i>PARTE 2: Desarrollo de los Métodos de Cálculo de MyoRec</i>	63
2.B.1 Introducción.....	63
2.B.2 Materiales y métodos.....	64
2.B.2.1 Detector de Picos	64
2.B.2.2 Calculador de Datos.....	67
2.B.2.3 Efectividad y Eficiencia de Cálculo	70
2.B.3 Resultados	71
2.B.3.1 Interfaz Gráfica	71
2.B.3.1.1 Aplicaciones Principales	71
2.B.3.1.2 Aplicaciones Complementarias	78
2.B.3.1.3 Aplicaciones de Configuración	79
2.B.3.2 Estructura de los Archivos de Datos.....	79
2.B.3.3 Rendimiento del Detector de Picos y del Calculador de Datos.....	80
2.B.3.4 Efectividad del Calculador de Datos.....	80
2.B.4 Discusión.....	81
<i>Capítulo 3: Uso de MyoRec en el Análisis del Comportamiento de Alimentación</i>	84
3.1 Introducción.....	84
3.2 Materiales y Métodos.....	85
3.2.1 Insectos.....	85
3.2.2 Alimentador Artificial y Sistema de Registro de Electromiografía	85
3.2.3 Soluciones.....	86
3.2.4 Análisis Estadístico.....	87
3.3 Resultados	87
3.3.1 Descripción de los Parámetros y Variables para el Análisis Cuantitativo del Comportamiento	87
3.3.2 Comportamiento de Alimentación hacia Sustancias de Diferente Valencia.....	88
3.4 Discusión.....	92
Discusión Final	98

Bibliografía.....	106
Apéndice A.....	117
A.1 Regresión Logística.....	118
A.2 Árboles de Decisión.....	121
A.3 Ensamblados.....	125
A.3.1 Votación	126
A.3.2 Bagging	126
A.3.3 Random Forest	128
A.3.4 Gradient Boosting.....	128
A.4 Máquinas de Soporte Vectorial (SVM)	130
A.5 Redes Neuronales Artificiales	141
A.5.1 Perceptrón Simple (SP).....	143
A.5.2 Perceptrón Multicapa (MLP)	145
A.5.3 Redes Neuronales Convolucionales (CNNs)	146
A.5.3.1 Principales Capas de una CNN.....	147
A.5.3.1.1 Capa de Convolución	148
A.5.3.1.2 Capa de Max/Average Pooling	150
A.5.3.1.3 Capa de Unidad Lineal Rectificada (ReLU)	151
A.5.3.1.4 Capa Dropout	152
A.5.3.1.5 Capa Flatten.....	152
A.5.3.1.6 Capa Fully Connected (FC).....	153
A.5.3.1.7 Capa de Clasificación	153
A.6 Redes Recurrentes	153
A.6.1 LSTM	154
APÉNDICE B.....	159
APÉNDICE C.....	162
C.1 Modelos de Machine Learning.....	162
C.2 Redes Neuronales Convolucionales	164

Introducción general

Todos los seres vivos necesitan alimentarse para vivir. Numerosos mecanismos fisiológicos y estrategias comportamentales han aparecido a lo largo de cientos de millones de años en diferentes organismos que habitan el planeta tierra, incluyendo a los insectos. Los mecanismos y comportamientos adoptados por los organismos en la búsqueda del alimento, que incluyen activación, detección y orientación son variados, aunque también pueden compartir similitudes que dependen de cada caso particular (Friend y Smith, 1977). Muchos comportamientos pueden ser aprendidos mediante la interacción del individuo con su entorno (Dukas, 2008) o ser completamente innatos, y con ello estar ajustados de forma iterativa mediante pasos evolutivos y asociados a una población de individuos (Kansaki, 1996; Liang y Brinkman, 2022).

La hematofagia, o el hábito de alimentarse de sangre, en los artrópodos, surgió al menos 5 veces de manera independiente en diversos grupos (Mans 2011, Ribeiro 1995; Marquart 2005). Sus antepasados posiblemente fueron depredadores (como los insectos redúvidos), saprófagos (como las garrapatas) o fitófagos (como los mosquitos y las moscas), siendo muy probable que en cada caso, algunos de sus respectivos ancestros hayan presentado características morfológicas y/o fisiológicas y/o comportamentales que facilitaron la aparición del hábito hematófago (Lehane, 2005). La sangre de los animales vertebrados, que sirve de alimento a los hematófagos, es rica en nutrientes que son necesarios para la supervivencia y la reproducción. Las adaptaciones específicas deberían incluir al menos sensores capaces de detectar la presencia de una fuente de sangre, piezas bucales adecuadas para perforar la piel, moléculas que inhiban la hemostasia, mecanismos involucrados en la digestión y absorción de nutrientes de la sangre, proteínas desintoxicantes para evitar efectos nocivos y excreción de moléculas desechables.

Sin embargo, acceder a la sangre contenida en un vaso sanguíneo bajo la piel de un vertebrado, implica una serie de obstáculos y riesgos que pueden provocar la muerte al insecto en caso de ser detectado. Además, existe la posibilidad de que el alimento contenga sustancias de baja palatabilidad, que incluso pueden resultar nocivas para el insecto al entrar en contacto

con ellas o luego de ingerirlas (Barrozo, 2019). La presencia de estos compuestos en la sangre del vertebrado puede ser una consecuencia de la herbivoría de plantas que producen metabolitos secundarios para defenderse de los depredadores. Afectan tanto a los insectos que se alimentan directamente de estas plantas como a los entomófagos que se alimentan de ellos, y a los vertebrados que al consumir las plantas incorporan circunstancial y temporalmente las moléculas nocivas en la sangre. En consecuencia, el mecanismo de detección de estas sustancias en insectos hematófagos puede haberse heredado de un sistema conservado evolutivamente en insectos fitófagos y entomófagos (Lehane, 2005).

Muchos insectos hematófagos son vectores de enfermedades que afectan al ser humano, como por ejemplo la malaria, la enfermedad del sueño, filariasis, leishmaniasis, dengue, tifus, entre otras. También provocan extensos daños a la ganadería debido a las enfermedades que provocan al ganado mediante la transmisión de parásitos (Lehane, 2005; WHO, 2023a; Ortega-Insaurralde y Barrozo, 2022; Cuthbert et al., 2023; Guhl et al., 2009; Petersen et al., 2016; Severson y Behura, 2016). En particular, los triatominos son los responsables de transmitir la enfermedad de Chagas al Hombre, que afecta a un sector importante de la población en América Latina (WHO, 2023b). La importancia sanitaria de estos animales es evidente, y *Rhodnius prolixus* ha sido utilizado desde mediados del siglo XX como un modelo en el estudio de la fisiología de los insectos (Nunes-da-Fonseca et al., 2017). La enfermedad de Chagas, al igual que otras enfermedades transmitidas por insectos hematófagos, puede prevenirse a través de múltiples enfoques, como por ejemplo el control vectorial (Gorla y Hashimoto, 2017). El control vectorial es especialmente importante debido a que la enfermedad solo puede curarse en una fase temprana con medicamentos altamente costosos (e.g. quimioterapia) que inducen efectos secundarios importantes (Pérez-Molina et al, 2020; Castro et al., 2015). Entender el comportamiento alimenticio de estos animales y sus mecanismos de interacción con el medio ambiente, y en particular, cuando entra en contacto con el hospedador, puede ayudar al desarrollo de métodos de control y monitoreo que disminuyan su interacción con el ser humano.

A continuación, se describe el comportamiento alimenticio de los insectos hematófagos, mencionaremos y discutiremos las diferentes etapas de la búsqueda de alimento, enfocándonos

principalmente en la interacción entre el insecto y el hospedador. Describiremos el sistema sensorial involucrado en la evaluación del alimento (*i.e.* sistema gustativo), describiremos la composición del alimento que ingieren (*i.e.* la sangre), y en particular algunos tipos de compuestos que pueden estar contenidos en él y que pueden afectar su palatabilidad. Mostraremos cómo se registra el comportamiento alimenticio y cuáles son algunas de las técnicas más importantes que existen hoy en día para analizarlo. Describiremos el modelo biológico que estudiamos en esta tesis, las características de su aparato bucal, el mecanismo que emplea para acceder y extraer el alimento de su hospedador, y la relevancia que tiene este trabajo en el ámbito académico y sanitario. Finalmente, describiremos las diferentes hipótesis propuestas y los objetivos que queremos alcanzar con la investigación.

I.1 Comportamiento de Búsqueda del Alimento

En los insectos hematófagos, la búsqueda y localización del hospedador se realiza a distancia utilizando claves olfativas, calor y humedad emitidas por sus hospedadores vertebrados, incluyendo a los humanos (Lehane, 2005; Barrozo y Lazzari, 2004a, b; Barrozo et al., 2004, Guerenstein y Guerin, 2001). Varios componentes comportamentales en ésta búsqueda por el alimento son comunes a diferentes insectos hematófagos, los cuales se han denominado: **activación, detección, orientación y contacto con el hospedador** (Lazzari, 2009; Lehane, 2005; Barrozo, 2019). La **activación** implica un incremento repentino de la actividad del insecto y la ejecución de movimientos no orientados dirigidos a la búsqueda de claves sensoriales emitidas por el hospedador. Está regulada por un control endógeno sincronizado con el período de sueño o baja actividad del hospedador. La **detección** se produce cuando el animal percibe claves o estímulos sensoriales producidos por la fuente de alimento, lo que incrementa el estado de alerta y produce comportamientos específicos del insecto en cuestión, como el levantamiento de las antenas y/o la suspensión instantánea del movimiento. Por lo tanto, la activación y la detección son eventos que suelen ocurrir a larga distancia entre el receptor (insecto) y el emisor (hospedador) en la que imperan principalmente las claves olfativas. La **orientación** tiene lugar cuando el animal cambia de rumbo y comienza a avanzar hacia el hospedador en una trayectoria más directa. A medida que el animal se acerca a su objetivo, otras claves emitidas por la fuente

cobran mayor relevancia. Los cambios de temperatura y humedad se incrementan, y la localización visual de la fuente de alimento es cada vez más fácil. Por lo tanto, la orientación es un evento que se manifiesta cuando existe una distancia media-larga entre el receptor y el emisor. Finalmente, el insecto se posa sobre la piel del hospedador y se produce la interacción o **contacto**, que está conformado por una etapa de *reconocimiento* y otra de *ingestión*. En la etapa de reconocimiento el animal utiliza órganos sensoriales externos ubicados en las antenas, piezas bucales externas y tarsos de las patas para evaluar la calidad del hospedador, mientras que, en la etapa de ingestión, inserta sus estiletes en el interior de la piel hasta encontrar un vaso sanguíneo, y la sangre ingerida es evaluada por órganos sensoriales internos.

I.2 Comportamientos durante la Interacción o Contacto con el Hospedador

I.2.1 Etapa de Reconocimiento

Una vez que el insecto encuentra al hospedador y establece contacto con la piel, emprende la búsqueda de un vaso sanguíneo para alimentarse de sangre. Para localizar un capilar sanguíneo el animal tiene dos opciones: perforar la piel en ocasiones múltiples hasta encontrar la fuente, o utilizar algún mecanismo especialmente adaptado para detectarla. Para picar azarosamente la piel no se necesita de ningún mecanismo especial, pero implica generalmente mayor tiempo de contacto y expone al insecto a una mayor probabilidad de ser detectado por el hospedador. Estudios realizados en insectos triatomíneos muestran que las zonas bajo la piel donde se sitúan los vasos sanguíneos, son ligeramente más calientes que las zonas circundantes, lo que permite al insecto hacer uso de termorreceptores ubicados en las antenas para localizar estos vasos (Ferreira *et al.*, 2007). Es gracias a esta detección que los insectos encuentran la fuente de calor (Flores y Lazzari, 1996). Luego de dicha detección térmica los insectos triatomíneos se dirigen de forma orientada hacia las zonas de la piel de mayor temperatura que suelen estar más irrigadas de sangre para picar. La humedad es otra clave que juega un papel clave en la orientación a corta distancia de los triatomíneos hacia el hospedador (Barrozo y Lazzari, 2004). Sin embargo, el calor sería la clave más relevante y necesaria para disparar en los insectos triatomíneos el despliegue o el estiramiento de su piezas bucales o proboscis hacia el sitio de picado (Flores y Lazzari, 1996). Sin embargo, la picadura no se producirá si el insecto detecta

sobre la piel del hospedador determinadas sustancias (Barrozo, 2019). El animal utiliza quimiorreceptores de contacto ubicados en distintas partes de su cuerpo para evaluar la presencia de sustancias aversivas en el sitio de picado, lo que representa la última instancia de evaluación de la etapa de reconocimiento antes de tomar la decisión de perforar la piel con sus piezas bucales (Barrozo, 2019).

I.2.2 Etapas de Ingestión

La etapa de ingestión comienza cuando el insecto inserta sus piezas bucales en la piel del hospedador en la búsqueda de un vaso sanguíneo. Se caracteriza por la necesidad del animal de succionar porciones de sangre para examinar y consecuentemente iniciar el proceso de alimentación. Estos dos momentos se denominan **Fase de Muestreo** y **Fase de Verdadera Ingestión** respectivamente (Barrozo, 2019).

I.2.3 Fase de Muestreo

Una vez que el insecto perfora la piel con sus mandíbulas, sus estiletes se extienden en el interior realizando movimientos laterales en búsqueda de un vaso sanguíneo para canular (**Figura 1**). Durante este proceso, conocido como **Fase de Muestreo** (Smith, 1970), el animal toma pequeñas cantidades de sangre del capilar (Friend y Smith, 1977). La detección de ciertas moléculas en el alimento succionado pueden desencadenar la ingestión sostenida y regular de sangre.

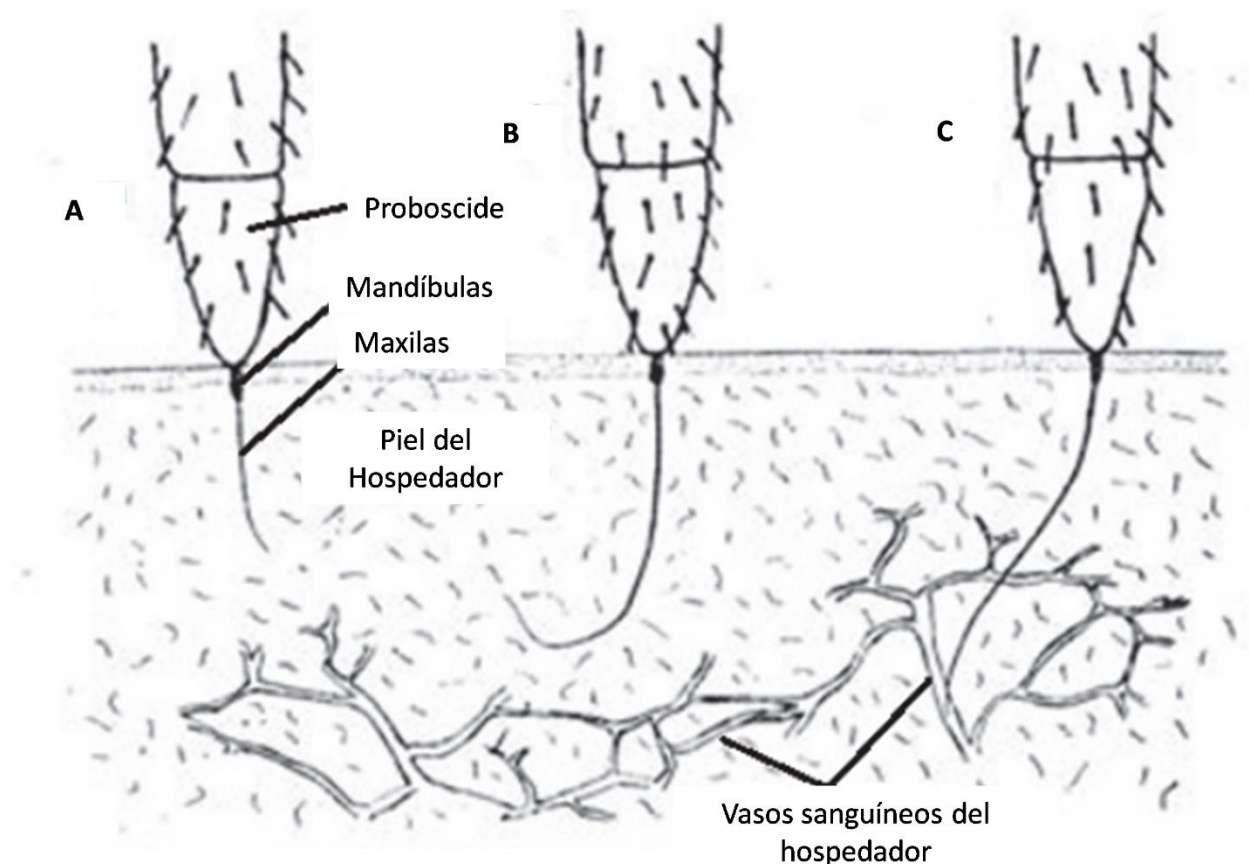


Figura I.1. Etapa de ingestión en un insecto triatomino. **A** – Daño de la piel con las mandíbulas e inserción de los estiletes o maxilas en la piel del hospedador. **B** – Movimientos laterales de los estiletes para buscar un vaso sanguíneo. **C** – Inserción de los estiletes en un vaso sanguíneo para dar inicio a la fase de muestreo y posteriormente a la fase de verdadera ingestión. Extraído de Araujo *et al.*, (2011).

I.2.4 Fase de Verdadera Ingestión

El proceso de succión sostenida de la sangre se denomina fase de verdadera ingestión, y se produce una vez que el insecto decide alimentarse tras la fase de muestreo (**Figura I.1**). En mosquitos, triatominos y otros hematófagos, un conjunto de músculos ubicados en la cabeza del animal, llamados bomba cibarial y faríngea, se relajan y contraen regularmente, generando una

presión negativa a través del canal de alimentación suficiente para posibilitar el bombeo o succión de la sangre (**Figura I.2**).



Figura I.2. Canal alimenticio, bomba cibarial y bomba faríngea del mosquito *Aedes albopictus*. Modificado de Kikuchi *et al.* (2018).

I.3 Sistema Gustativo: Reconocimiento del Alimento

El sabor del alimento proporciona a un animal información acerca de la calidad potencialmente nutritiva o tóxica del mismo. El sistema sensorial que confiere a los animales esta capacidad de detección es el sistema gustativo, que en los insectos se denomina también quimiorrecepción de contacto. Este sistema sensorial está altamente especializado en la detección de sustancias de elevada palatabilidad o apetitivas (*e.g.*, azúcares, baja concentración de sales y aminoácidos) y de baja palatabilidad o aversivas (*e.g.*, sustancias amargas, ácidas y alta concentración de sales). Cuando los insectos entran en contacto con la piel del hospedador, el sistema gustativo o quimiorrecepción de contacto provee la capa de información decisiva para que el insecto decida picar o no, alimentarse o no (Barrozo, 2019).

I.3.1 Sistema Gustativo o Quimiorrecepción de Contacto en Insectos

El sistema gustativo o quimiorrecepción de contacto en los insectos está conformado por un conjunto de elementos sensoriales que se ubican en diferentes **órganos** del cuerpo del insecto, especialmente en las antenas y las piezas bucales, aunque con frecuencia también se hallan en las patas, el aparato reproductor e incluso en las alas en menores cantidades (Liman *et al.*, 2014). Estos elementos están formados por neuronas sensoriales envainadas por una capa cuticular en forma de pelo que presenta un poro en la parte apical. Tales estructuras se conocen como **sensilias gustativas** o **quimiorreceptores de contacto**. Las neuronas sensoriales gustativas se activan cuando entran en contacto con compuestos químicos presentes en el sustrato en estado líquido o sólido (**Figura I.3**). Las moléculas entran a través del poro apical de la sensilia, y, en algunos casos, con ayuda de proteínas transportadoras, se acoplan a receptores, *i.e.*, proteínas transmembranales ubicadas en las dendritas de estas neuronas. Cuando esto sucede, se produce una cascada de señalización con la presencia de segundos mensajeros, que deriva en apertura o cierre de canales iónicos produciendo un cambio de potencial de membrana y la producción de potenciales de acción o señales eléctricas que se extienden hasta el cerebro del insecto. En particular, las neuronas sensoriales gustativas (NSGs) proyectan hacia neuropilos del sistema nervioso central donde la información será integrada y procesada (Koul, 2005).

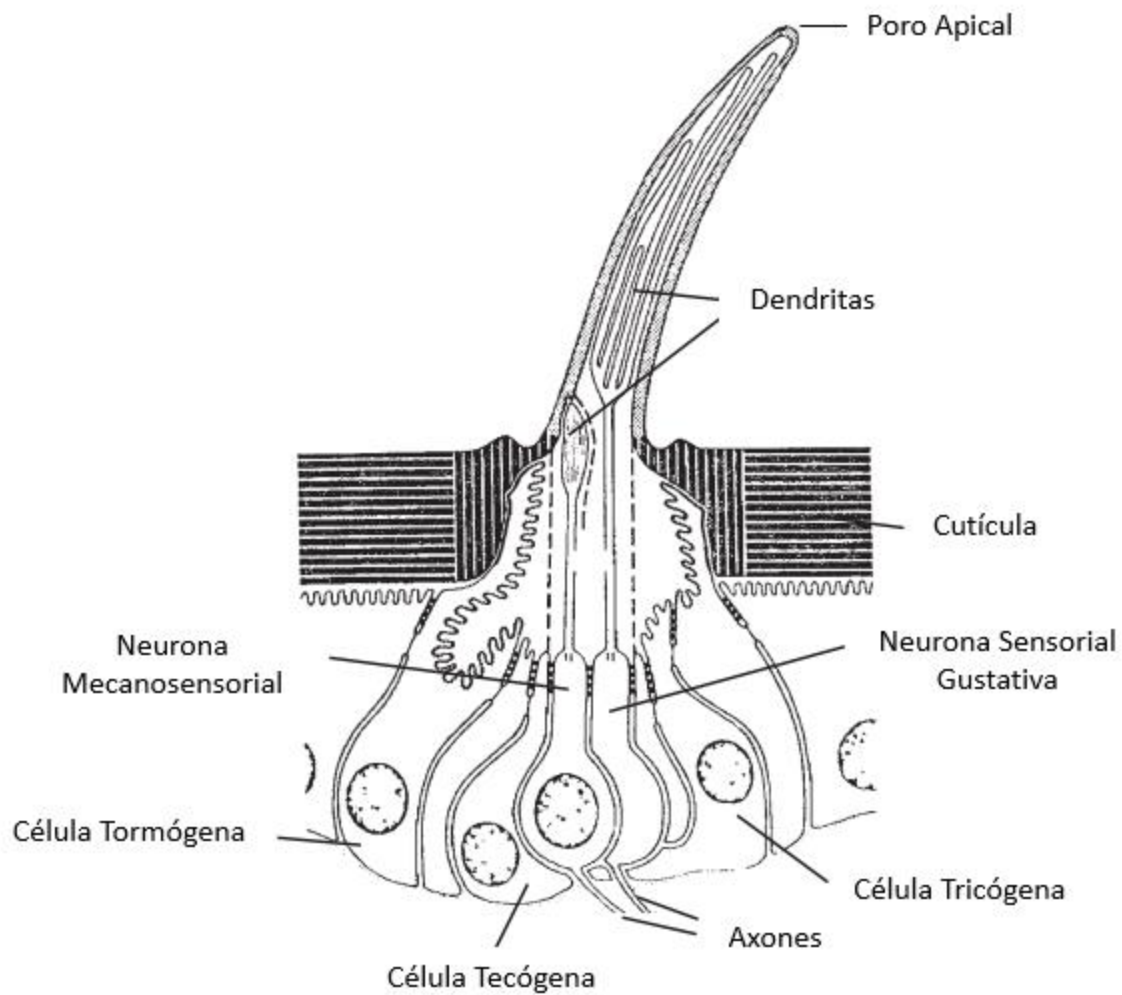


Figura I.3. Estructura de una sensilia gustativa. Extraído de (Klowden, 2007).

Los quimiorreceptores de contacto pueden ser externos o internos. Los quimiorreceptores externos tienen la función de detectar la presencia de moléculas presentes sobre el sustrato o la piel de un hospedador (**Figura I.4**). Los quimiorreceptores internos, por su parte, están ubicados en la faringe, y les permiten sensor la calidad y composición química del alimento ingerido (**Figura I.5**).

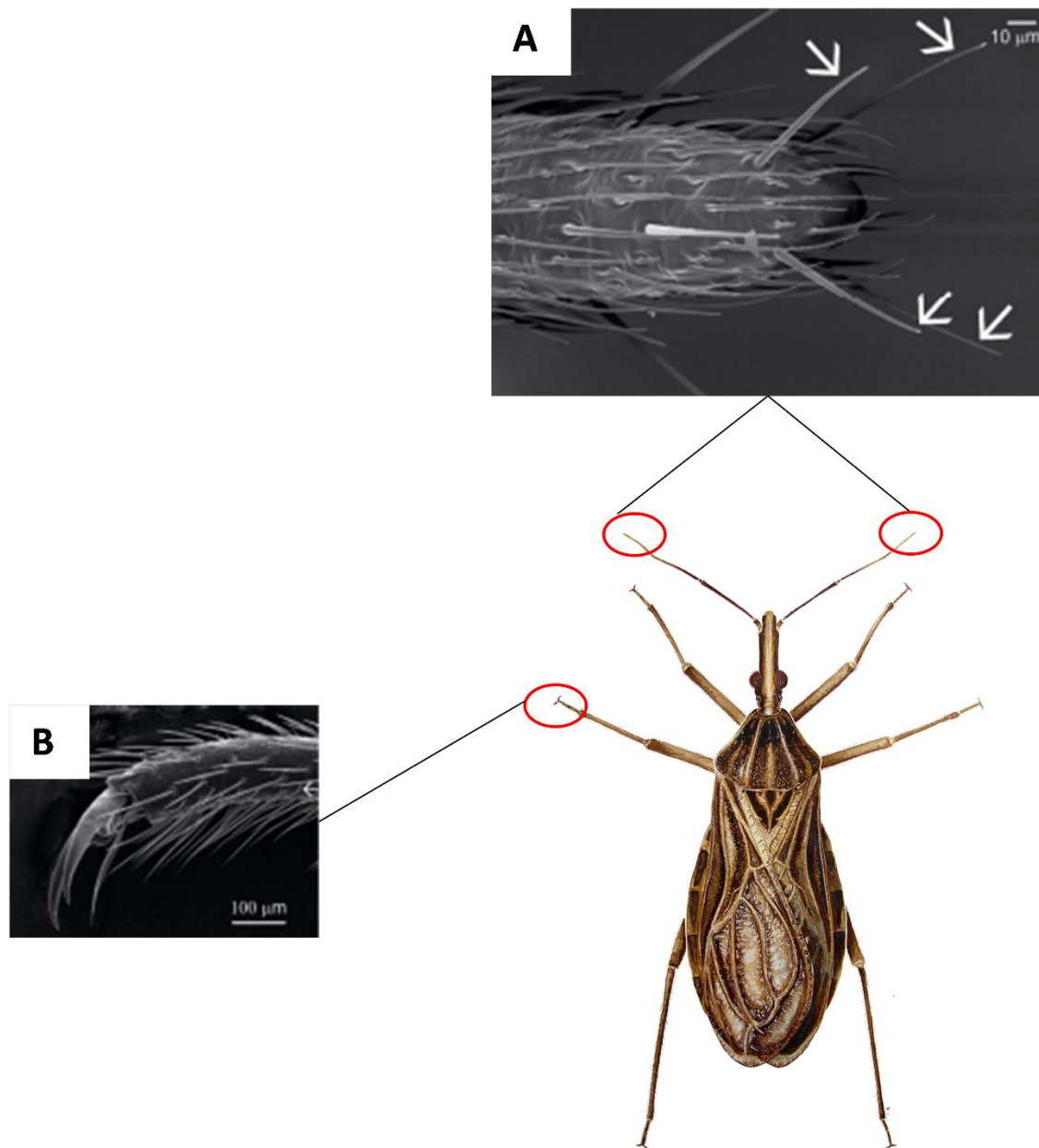


Figura 1.4. Quimiorreceptores de contacto externos en un adulto de *R. prolixus*. A – Quimiorreceptores de contacto del segmento distal de la antena (indicadas con flechas blancas). B – Múltiples sensilias del tarso de *R. prolixus*. Figura del insecto extraída de www.irenebehrens.com. Fotografías de microscopía electrónica de barrido obtenidas de (Barrozo *et al.*, 2017).

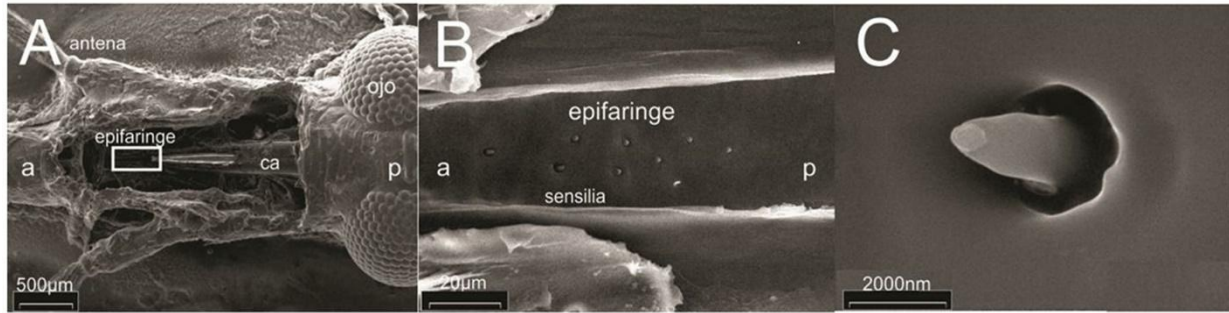


Figura I.5. Quimiorreceptores de contacto internos en un adulto de *R. prolixus* vistos a través de imágenes de microscopía electrónica de barrido. **A** - Vista superior de cabeza y de la faringe. **B** - Quimiorreceptores de contacto en la epifaringe de *R. prolixus*. **C** - Vista de un quimiorreceptor de contacto con un poro apical. Extraído de (Pontes *et al.*, 2014).

El papel que desempeña la gustación o quimiorrecepción de contacto en el reconocimiento de un hospedador o fuente de alimento ha sido muy poco explorado en insectos hematófagos (Barrozo, 2019; Benton, 2017). En la actualidad sabemos que, durante el largo y complejo proceso de alimentación, la quimiorrecepción de contacto cumple su rol en dos etapas: en una primera instancia en la evaluación de la piel del hospedador mediante quimiorreceptores externos (**Figura I.4**), y en una segunda instancia cuando la sangre es ingerida y evaluada por los quimiorreceptores internos (Pontes *et al.*, 2014; Dennis *et al.*, 2019; Jové *et al.*, 2020) (**Figura I.5**). Ambas etapas son relevantes para el avance y conclusión de la alimentación de estos insectos: durante la evaluación externa de la piel el insecto decidirá si pica o no, mientras que durante la evaluación interna de lo ingerido decidirá si come o no (Barrozo, 2019).

I.4 El Alimento de los Insectos Hematófagos: la Sangre

La sangre de los organismos vertebrados está compuesta por una gran diversidad de moléculas que estimulan la alimentación de los insectos hematófagos. Sales inorgánicas (NaCl, KCl y NaHCO₃), glucosa, glutatión reducido, nucleótidos purinérgicos y albúmina, han resultado ser factores críticos en la aceptación del alimento en diferentes insectos hematófagos (Friend y Smith, 1977). Nucleótidos de adenosina (ATP, ADP), y otros nucleótidos (*e.g.*, GTP y GDP), son

liberados por eritrocitos y plaquetas en diferentes condiciones fisiológicas y han demostrado tener un poder fagoestimulante en varias especies de insectos hematófagos (Galun y Margalit, 1969; Friend y Smith, 1975; Friend y Smith, 1977; Galun y Kabayo, 1988). Es así, que el ATP y el ADP son los nucleótidos con mayor efecto fagoestimulante en diferentes especies de mosquitos, triatominos, moscas del establo, chinches de cama, mosca tse-tse, tábanos, pulgas de las ratas, moscardón azul (Lukenge et al 2022).

I.5 Registro del Comportamiento de Alimentación

Como se mencionó más arriba, los eventos de alimentación incluyen insertar las piezas bucales, comenzar a succionar y muestrear el alimento (**fase de muestreo**), iniciar la verdadera ingestión (**fase de verdadera ingestión**) y dejar de alimentarse. Aunque la secuencia en esta cadena de eventos que definen la alimentación es similar entre individuos, pueden existir ciertas variaciones interindividuales en factores tales como en la duración de cada uno de los eventos mencionados. El comportamiento alimenticio en insectos succionadores (*e.g.*, áfidos, vinchucas y mosquitos) ha sido estudiado históricamente mediante tres métodos: microscopía intravital, registros de penetración eléctrica o electropenetrografía (EPGs), y electromiogramas (EMGs) (Roberto *et al.*, 2017).

I.5.1 Microscopía Intravital

La microscopía intravital comenzó a utilizarse a finales de los años 30, cuando Gordon y Lumsden (1939) caracterizaron la ingesta de sangre en mosquitos mientras se alimentaban con ranas. Esta técnica consiste en la observación directa de los órganos involucrados en la alimentación por microscopía mientras ocurre la alimentación (**Figura 6 y 7**). Gracias a esta técnica se diferenciaron dos tipos de mecanismos de alimentación: cuando los insectos ingieren directamente de un vaso sanguíneo (*i.e.*, solenofagia) y cuando lo hacen de lagunas de sangre que se producen luego de que el insecto muerde la piel (*i.e.*, telmofagia) (Griffith y Gordon, 1952; Lavoipierre, 1965). La aparición de marcadores fluorescentes y técnicas de procesamiento de imágenes hacen que actualmente la técnica sea muy eficaz para describir los procesos

involucrados en la alimentación de artrópodos hematófagos (Choumet *et al.*, 2012; Soares *et al.*, 2014).

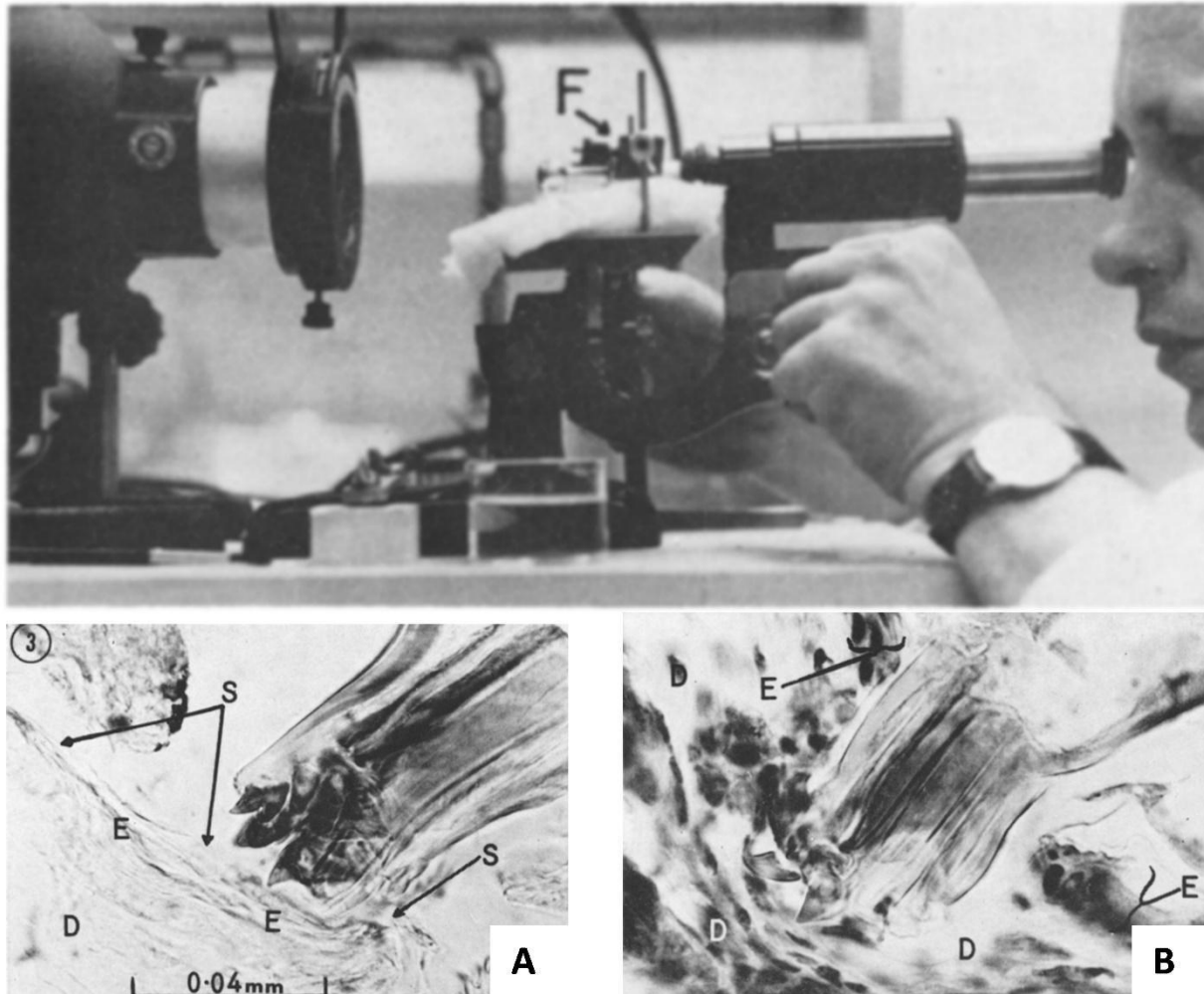


Figura I.6. Microscopio utilizado para observar la penetración de la probóscide de un piojo en la oreja de un ratón. F muestra la cámara de alimentación acoplada al escenario sobre el que se coloca el ratón. **A** – El extremo de la probóscide descansa sobre la superficie de la piel y las partes bucales se encuentran en proceso de eversión. **B** – La probóscide ha penetrado la piel y las partes bucales han sido evertidas. Extraído de Lavoipierre (1967).

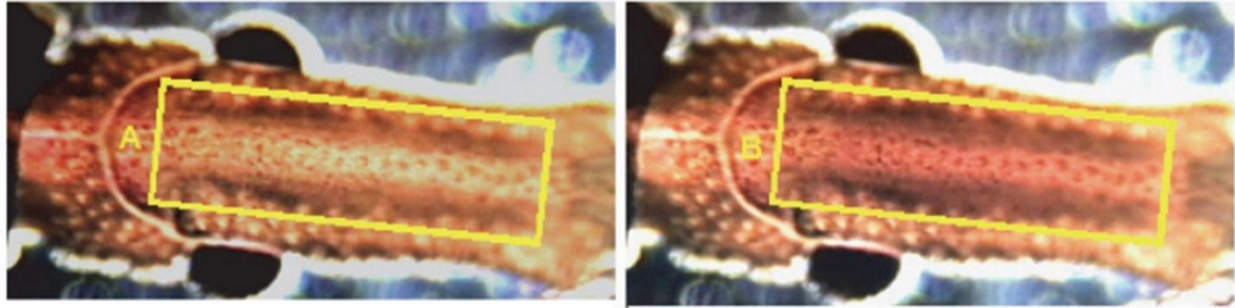


Figura I.7. Microscopía intravital utilizada para observar el flujo de sangre a través del canal alimenticio de ninfas de *R. prolixus*. **A** – El canal alimenticio se encuentra vacío. **B** – El canal alimenticio se llena de sangre. Extraído de Araujo *et al.* (2011).

I.5.2 Electropenetrografía

La electropenetrografía (EPG) permite detectar cambios en la resistencia eléctrica existente entre el insecto y su fuente de alimento (McLean y Kinsey, 1964). En el tórax del animal se conecta o pega un electrodo (*e.g.*, alambre delgado de oro) y luego se coloca el insecto en una planta que se encuentra conectada a una fuente de voltaje. Cuando los estiletes se insertan en la planta el circuito se cierra y el voltaje constante aplicado a la planta se vuelve variable a medida que pasa a través del insecto. Los voltajes variables que se muestran a lo largo del tiempo producen formas de onda que representan comportamientos del movimiento de los estiletes en la planta. Este sistema se desarrolló específicamente para estudiar el comportamiento alimenticio de hemípteros fitófagos, como los áfidos, y posteriormente se extendió a otros insectos de importancia agroeconómica (McLean y Kinsey, 1964; Backus *et al.*, 2019). La técnica también se ha adaptado para el uso en insectos hematófagos (Wayadande *et al.*, 2019).

I.5.3 Electromiografía

Smith y Friend (1970) estandarizaron la técnica EPG para su uso en insectos triatominos. Smith (1979) descubrió que podía registrar la actividad eléctrica generada por el insecto en ausencia de la batería o de la aplicación de voltaje en la fuente de alimento. El electrodo (*e.g.*, alambre de plata) se coloca también en el tórax del insecto y el otro electrodo en la solución

alimenticia (Araujo *et al.*, 2011). A esta variante de la técnica de EPG se le conoce como electromiografía y los resultados se visualizan a través de registros conocidos como electromiogramas (EMGs). Los EMGs son similares a los EPGs, pero se diferencian en que la señal que representan carece de una componente producida por la fuente de voltaje. Esta técnica se ha utilizado en triatominos durante décadas para revelar aspectos del comportamiento de alimentación que suceden en la interfaz insecto-hospedador. Aunque también ha sido utilizada en insectos no hematófagos como hormigas (Falibene y Josens, 2008; Lois-Milevicich *et al.*, 2020). Anteriormente se mencionó que los triatominos se alimentan gracias a la fuerza de succión de dos conjuntos de músculos, localizados en la parte anterior de la cabeza y que se contraen y relajan de forma antifásica: la bomba cibarial y la bomba faríngea (**Figura 8**). La señal producida por estas bombas es fácilmente registrable y cuantificable a través de los EMGs. En consecuencia, la electromiografía ha permitido establecer varios fundamentos del proceso de alimentación y el funcionamiento de las bombas (Smith y Friend 1971; Smith, 1979; Bennet-Clark 1963a, b).

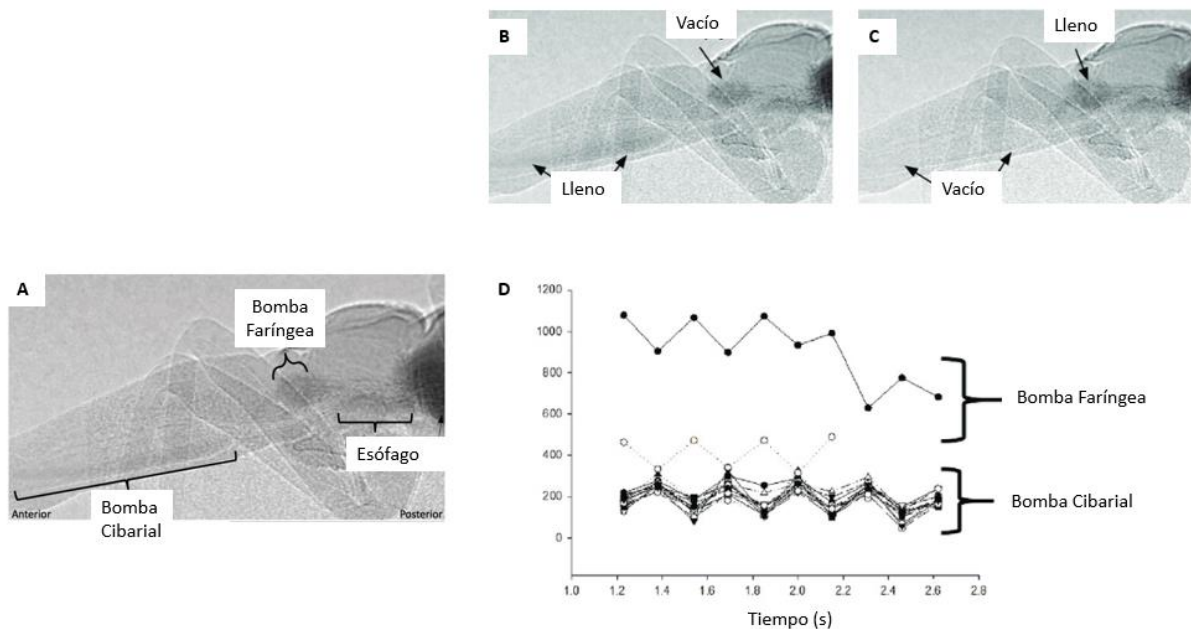


Figura 1.8. Bomba cibarial y faríngea de *R. prolixus*. **A** – Imagen de rayos X con vista sagital de la cabeza de una ninfa de *R. prolixus* durante su alimentación. **B - C** – Imágenes de rayos X que muestran la actividad de la bomba cibarial y faríngea durante la alimentación con sangre. **D** –

Actividad de la bomba cibarial y faríngea donde se observa que ambos mecanismos se sincronizan de forma antifásica. Extraído de Lahondere *et al.* (2017).

I.6 Patrones de Onda en EPGs y EMGs

Para extraer información de los EPGs y/o EMGs deben caracterizarse los patrones de onda asociados a comportamientos alimenticios de interés. Su identificación y asociación a las diferentes fases del proceso de alimentación es una tarea que se ha completado en diferentes especies de insectos. En áfidos, una serie de patrones particulares fueron caracterizados a mediados de la década de los 80' en EPGs. Estos incluyen la penetración y no penetración de los estiletes en el espacio intercelular de la planta, la salivación de la vaina del estilete, perforaciones celulares, ingestión de savia de xilema y floema, entre otros. La caracterización de estas señales contribuyó a extender notablemente el uso de la técnica en esos y otros insectos (Giordanengo, 2014; Wayadance, 2019).

En triatominos, se han identificado patrones de onda para algunos comportamientos como la fase de muestreo, el inicio de la fase de ingestión y el final del proceso de alimentación. Esto permitió definir algunos parámetros como, el número de picaduras, la tasa de ingestión, el tiempo de contacto con la fuente de alimento, la frecuencia de bombeo, entre otros (Araujo *et al.*, 2011; Pontes *et al.*, 2017).

I.7 El Modelo de Estudio de esta Tesis

Los triatominos son los vectores biológicos del protozoario flagelado *Trypanosoma cruzi*, agente causal de la enfermedad de Chagas o tripanosomiasis americana. La enfermedad puede ser transmitida por los triatominos al Hombre y a otros animales vertebrados, incluyendo aquellos que son domésticos (de Oliveira *et al.*, 2018). Es endémica en 21 países del continente americano y se estima que cerca de 100 millones de personas están en riesgo de infectarse, unos 8 millones están infectadas y 56.000 nuevos casos por año produciendo aproximadamente 12.000 muertes anuales. En la Argentina, existen cerca de 2 millones de infectados (WHO, 2020).

Según la Organización Mundial de la Salud, en el mundo se registran alrededor de 700.000 defunciones por año por enfermedades transmitidas por insectos vectores (WHO, 2020). La lucha contra la transmisión vectorial es multimodal e incluye la incorporación de hábitos culturales, aplicación de insecticidas y el uso de repelentes que eviten que seamos encontrados y/o picados por los insectos (Dell’Arciprete et al., 2014; Nava-Doctor et al., 2021; Wilson et al., 2020; Alzogaray y Zerba, 2017; Ramirez et al., 2020). En la generación de este último escudo de protección constituido por los repelentes, el panorama completo de cómo los insectos hematófagos reconocen y deciden alimentarse sobre un hospedador potencial es de vital importancia, dado que, una profunda comprensión sobre los factores que afectan este proceso esencial podría ayudarnos a encontrar nuevas estrategias de protección personal con la finalidad de reducir las interacciones entre los vectores y el ser humano.

Rhodnius prolixus es probablemente la especie de triatominos mejor estudiada por su importancia sanitaria y porque ha servido desde comienzos del siglo pasado como un insecto modelo para una amplia variedad de disciplinas, incluyendo bioquímica, fisiología, biología molecular, ecología, toxicología, entre otras (Wigglesworth 1934; Friend y Smith, 1977; Barrozo et al., 2017; Schilman, 2016; Alzogaray y Zerba, 2017; Schofield y Diotaiuti, 1999, Mesquita et al., 2015, entre otros).

I.7.1 Aparato Bucal de *Rhodnius prolixus*

El aparato bucal en los insectos triatominos y en particular, en *R. prolixus*, consiste en un labio o probóscide con tres segmentos que posee en su interior dos estructuras serradas llamadas mandíbulas, y otras dos de mayor longitud y flexibilidad denominadas maxilas. El insecto utiliza las mandíbulas para perforar la epidermis y extender sus maxilas a través de la dermis, moviéndolas lateralmente hasta insertarlas en una vena o arteriola (**Figura 9**). Cada maxila se articula junto a la otra generando dos canales para el transporte de fluidos; el canal alimentario, utilizado para la absorción de sangre, y el canal salival, que conduce saliva a la zona de alimentación y posee un diámetro más reducido. La succión de la sangre está mediada por la acción de las bombas cibarial y faríngea, que permiten succionar la sangre a través del canal alimentario y dirigirla al sistema digestivo.

La articulación de todos estos elementos, especialmente adaptados para extraer las porciones de alimento que el animal necesita, permiten neutralizar parcialmente la respuesta nociceptiva, hemostática, inflamatoria e inmune del animal vertebrado durante la picadura e ingestión del insecto (Ribeiro y Francischetti, 2003; Soares *et al.*, 2006). De esta manera, el insecto facilita la extracción del alimento y reduce el riesgo de ser detectado y eliminado por el hospedador.

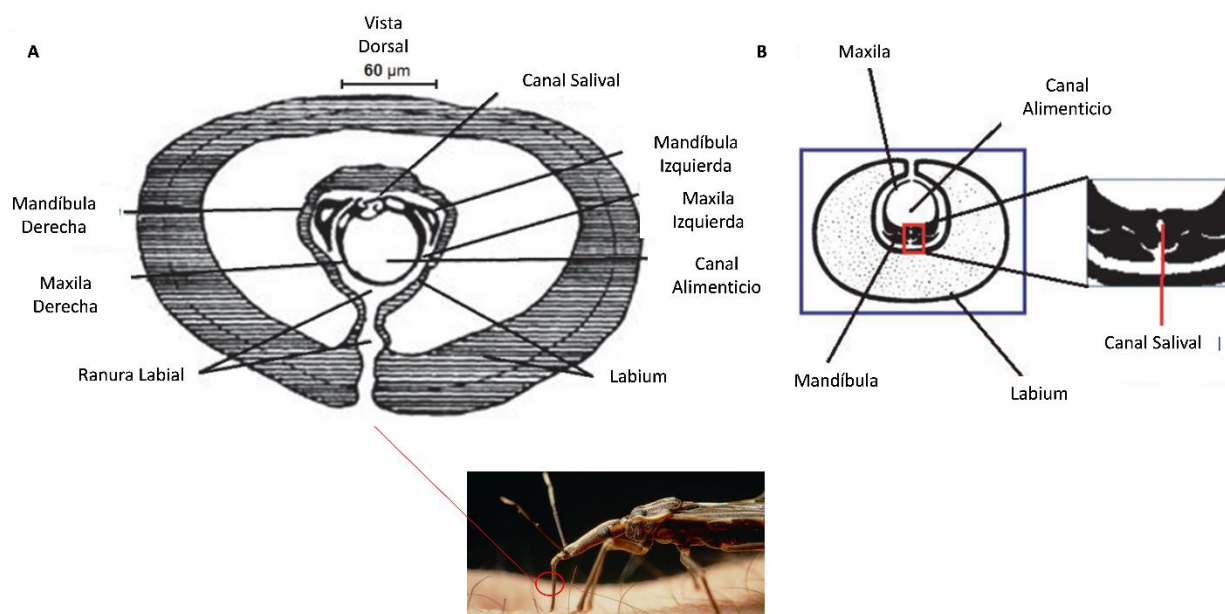


Figura 1.9. Aparato bucal de *Rhodnius prolixus* y *Cimex lecturalis*. **A** – Sección transversal del aparato bucal cerca de la base de la probóscide en una ninfa de *R. prolixus*. **B** - Sección transversal del aparato bucal cerca de la base de la probóscide en una ninfa de *C. lecturalis*. Modificado de (Araujo *et al.*, 2011).

1.7.2 Estructura del Principal Músculo Suctor de *Rhodnius prolixus*: la Bomba Cibarial

La bomba cibarial es un complejo de músculos ubicado en la cabeza del insecto que se contraen y relajan de forma periódica para controlar el flujo de sangre a través del canal alimenticio. La actividad de la bomba se produce durante las fases de muestreo y de verdadera

ingestión. En una ninfa de quinto estadio, la bomba está formada por una cámara dispuesta ventralmente en forma de U con 3,5 mm de largo y 0,28 mm de diámetro (Araujo *et al.*, 2011). Posee dos ligamentos a los lados de mayor elasticidad que se mueven junto a los músculos en forma de pistón (**Figura 10**), y el diámetro apical de la cámara una vez abierta es de 10 μm . La cámara está unida a la parte inferior de la cabeza mediante tejido intersticial, mientras que los pistones se unen a la parte superior mediante dos pares de músculos con forma de V. El primer par de músculos es considerablemente más largo y se ubica en la zona anterior de la cámara, mientras que el segundo es unas siete veces más corto y se encuentra en la parte distal (**Figura 10**) (Bennet-Clark, 1963b). La cámara que forma parte del canal alimentario se llena de sangre cuando se contraen estos músculos y se vacía cuando estos se relajan y los pistones se retraen (Snodgrass, 1935), asistidos por la válvula pre-cibarial y sus músculos asociados (Backus y McLean, 1982; McLean y Kinsey, 1964). Una ninfa de quinto estadio de *R. prolixus* puede absorber hasta 300 μl de sangre en 5 minutos, lo que significa que en la cámara se puede generar una presión negativa entre una y dos atmósferas. La frecuencia de bombeo puede variar entre 6 Hz (cuando el animal bombea la sangre regularmente y de forma estable) y 1 Hz (cuando el insecto tiene problemas para absorber el alimento) durante la fase de verdadera ingestión (Araujo *et al.*, 2011).

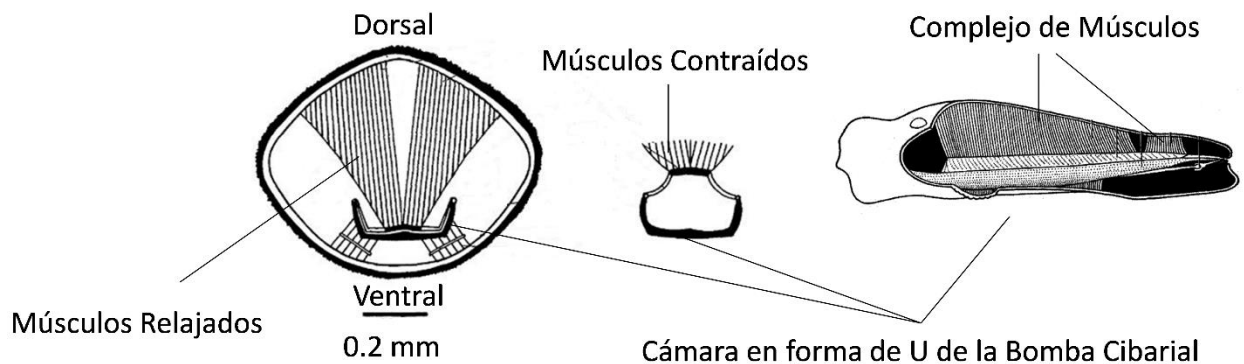


Figura I.10. Sección transversal y longitudinal de la cabeza de una ninfa de quinto estadio de *Rhodnius prolixus* que muestra la forma de U de la bomba cibarial y el conjunto de músculos asociados contraídos y relajados. Modificado de Araujo *et al.* (2011).

I.8 Compuestos Presentes en el Alimento: *Apetitivos versus Aversivos*

Como se mencionó previamente, entre los compuestos apetitivos presentes en la sangre de los hospedadores vertebrados se encuentra el ATP (Friend y Smith 1977, Romero y Schal, 2014). Muchos insectos hematófagos son altamente sensibles al ATP (*e.g.*, triatominos: *R. prolixus*, mosquitos: *Aedes aegypti*, pulgas, moscas tse-tse, chinches de cama, etc.), mientras que otros responden mejor al difosfato de adenosina o ADP (*e.g.*, moscas del establo, moscas negras, mosquitos: *Culex pipiens*, *Culiseta inornata*) (Barrozo, 2019). Por otra parte, existen moléculas de sabor amargo para el ser humano tales como la quinina, cafeína y la quinidina, que tienen un efecto anti-alimentario en los mosquitos *Anopheles gambiae* y *Ae. aegypti* (Ignell *et al.*, 2010; Kessler *et al.*, 2013; 2014; Dennis *et al.*, 2019).

Utilizando un alimentador artificial se demostró que la presencia de compuestos amargos (quinina y cafeína) o de sales inorgánicas en altas concentraciones (NaCl, KCl) sobre la superficie de un alimentador artificial donde se posa el insecto, reducen la alimentación de *R. prolixus* (Pontes *et al.*, 2014; 2022). En este mismo insecto también se demostró que la detección, a través de quimiorreceptores de contacto internos, de quinina, cafeína y de altas concentraciones de sales inorgánicas, también inhiben la alimentación (Pontes *et al.*, 2014; 2017).

I.9 Relevancia de esta Tesis

Desde un punto de vista antropocéntrico, la hematofagia asociada al ser humano abrió la puerta a la transmisión de enfermedades infecciosas. Según la OMS, las enfermedades transmitidas por vectores representan el 17 % de las enfermedades infecciosas, que afectan a cientos de millones de personas en todo el mundo, y se indica que todas son prevenibles utilizando medidas de prevención fundamentadas (WHO, 2017). El estudio de la contribución del sistema gustativo en el reconocimiento y evaluación del hospedador ha sido relegado, ya que se ha priorizado el estudio de sentidos para la detección de estímulos a larga distancia, como el sentido del olfato y el térmico (Pontes *et al.*, 2014). La evaluación del alimento previa a su ingestión es vital para todos los organismos, ya que la ingestión de compuestos nocivos puede producir efectos tóxicos en los individuos y con ello tener serias consecuencias fisiológicas en

desmedro de su salud (Muñoz *et al.*, 2020; Pontes *et al.*, 2014; Pontes *et al.*, 2017). De la misma manera, la ingestión de compuestos carentes del adecuado valor nutricional también pueden impactar negativamente en los organismos, disminuyendo sus posibilidades de supervivencia y éxito reproductivo (Muñoz *et al.*, 2020; Ladino, 2021; Glendinning *et al.*, 2001; Glendinning, 2007). El sistema gustativo es un sistema sensorial que le permite a los animales discriminar entre estas situaciones, y desencadena respuestas binarias: picar o no picar, comer o no comer (Barrozo, 2019). Dado este tipo de respuestas mayormente estereotipadas de aceptación o rechazo, resulta relevante poder identificar los compuestos que las producen y caracterizar los comportamientos generados. Este tipo de estudios en insectos vectores como las vinchucas entre otros, puede proveernos de nuevas herramientas que complementen o superen las actuales estrategias de control de la transmisión de enfermedades infecciosas transmitidas por insectos vectores.

I.10 Objetivos de esta Tesis

I.10.1 Objetivo General

Estudiar el comportamiento de alimentación de la vinchuca *Rhodnius prolixus* frente a distintas fuentes de alimento.

I.10.2 Objetivos Específicos

1. Registrar la actividad de los músculos de succión de los insectos durante la alimentación y caracterizar el comportamiento de alimentación a través del reconocimiento de patrones típicos de la actividad frente a compuestos apetitivos y aversivos.
2. Desarrollar un sistema de reconocimiento automático de patrones para la caracterización del comportamiento de alimentación en insectos triatominos.
3. Desarrollar un programa computacional para el registro y el análisis del comportamiento de alimentación de las vinchucas.
4. Implementar dicho programa en el análisis del comportamiento de alimentación de las vinchucas.

I.11 Hipótesis

I.11.1 Hipótesis 1

La decisión de un insecto de alimentarse está controlada por la detección de determinadas moléculas, las cuales desencadenarán respuestas apetitivas, *i.e.* aceptación del alimento, o respuestas aversivas, *i.e.* de evitación de la ingestión. Ambos tipos de respuestas producirán patrones identificables y característicos de cada condición alimenticia producto de la activación de los músculos involucrados.

I.11.2 Hipótesis 2

Si las etapas de la alimentación en insectos hematófagos son moduladas por moléculas presentes en el alimento que pueden ser detectadas sensorialmente, entonces es posible caracterizar la complejidad del comportamiento alimenticio a través de un modelo matemático.

I.11.3 Hipótesis 3

Diferentes modelos matemáticos son capaces de detectar patrones en series o señales temporales de cualquier tipo, entonces es posible elaborar un identificador de patrones que reconozca y delimite automáticamente los comportamientos de alimentación.

Capítulo 1: Caracterización del Comportamiento de Alimentación de *Rhodnius prolixus* a Compuestos Apetitivos y Aversivos

1.1 Introducción

El comportamiento de alimentación de los insectos triatominos está conformado por dos etapas importantes: prueba o muestreo del alimento e ingestión. De esta manera el insecto ingiere una pequeña porción de sangre, y realiza una prueba o evaluación gustativa de la misma. Si la sangre reúne las características necesarias, el insecto continúa con la verdadera ingestión (Barrozo, 2019).

La detección de compuestos apetitivos o fagoestimulantes en la sangre provocará que el insecto se alimente; en el caso que se detecten moléculas aversivas o anti-apetitivas el insecto abandonará la fuente en busca de otro sitio para picar (Smith y Friend, 1971). Se sabe que para iniciar la ingestión *R. prolixus* requiere detectar ATP y una concentración de NaCl similar a la presente en el plasma de la sangre. Esto ha permitido la creación de soluciones artificiales alimenticias basadas en estas moléculas, para ser usadas en experimentos con alimentadores artificiales desde hace al menos 50 años (Friend, 1965). Las soluciones salinas con ATP, han demostrado tener una efectividad para activar la alimentación en triatominos similar a la obtenida con la sangre de un hospedador vivo (Smith y Friend, 1970). El agregado, en la solución a ingerir, de moléculas amargas para el ser humano (cafeína, quinina, berberina, salicina), o de sales inorgánicas (NaCl, KCl) en una concentración mayor con respecto al plasma de la sangre de un vertebrado ($>0,2M$), inhiben la ingestión (Pontes *et al.*, 2014; 2017; Cano *et al.*, 2017). En la detección de moléculas tanto apetitivas como aversivas están involucrados los quimiorreceptores de contacto presentes en la faringe de las vinchucas, el órgano faríngeo (Barth, 1952; Barrozo *et al.*, 2017). El efecto inhibitor de las sustancias amargas (en un rango muy amplio de concentraciones) supera el efecto fagoestimulante del ATP y del NaCl (Asparch *et*

al., 2016; Pontes *et al.*, 2014). Esta inhibición podría ocurrir a través de la actividad directa de neuronas sensibles a compuestos amargos en el órgano faríngeo o a través de la inhibición directa de neuronas sensibles al ATP (French *et al.*, 2015).

Caracterizar el comportamiento de alimentación en estos insectos es relevante al momento de poder comprender y conocer cuáles son las moléculas que favorecen la ingestión y cuáles por el contrario la inhiben. Con tal propósito, se utiliza la técnica de registro de EMGs, mediante la cual es posible registrar la actividad de los músculos suctores localizados en la cabeza de los insectos durante la alimentación (*e.g.*, Pontes *et al.*, 2017; Guarneri *et al.*, 2000; Sant'Anna *et al.*, 2001).

El objetivo de este capítulo es registrar la actividad de los músculos de succión de los insectos durante la alimentación y caracterizar el comportamiento de alimentación a través del reconocimiento de patrones típicos de la actividad frente a compuestos apetitivos (ATP/NaCl) y aversivos (ATP/NaCl/cafeína). Poder definir las etapas de alimentación, las variables y los parámetros necesarios para caracterizar el comportamiento de alimentación nos permitirá sentar las bases para la creación de un programa computacional que permita realizar los análisis de forma automática, como se verá en los Capítulos 2 y 3.

1.2 Materiales y Métodos

1.2.1 Insectos

Se utilizaron ninfas del quinto estadio de *R. prolixus*. Los animales se criaron bajo condiciones controladas de temperatura (28 ± 1 °C) y fotoperíodo (12h de luz, 12h de oscuridad) y de humedad relativa ambiental (30 - 60 %). Luego de la eclosión, los insectos se alimentaron con sangre de gallina hasta alcanzar el 5° estadio ninfal. Las ninfas de 5° estadio fueron separadas luego de la ecdisis del resto de los insectos y mantenidas sin alimentar por 15 ± 2 días, momento en el que fueron asignados a un grupo experimental. Los animales fueron manipulados de acuerdo con las normas de bioseguridad de la Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.

1.2.2 Soluciones

Para los experimentos se utilizaron dos soluciones con conocido efecto en el grado de aceptación o rechazo en *R. prolixus* (Pontes *et al.*, 2014). Como solución apetitiva se utilizó 1 mM ATP (trifosfato de adenosina, CAS [56-65-5]) en 0,15 M NaCl (cloruro de sodio, CAS [7647-14-5]). Como solución aversiva, se utilizó 5 mM cafeína (cafeína anhidra, CAS [58-08-2]) (**Figura 1.1**). Esta última fue preparada en 0.15 M NaCl y 1 mM ATP para asegurar la motivación alimentaria de los insectos (Pontes *et al.*, 2014). Todas las soluciones se prepararon inmediatamente antes de los experimentos y se utilizó agua destilada como solvente.

1.2.3 Alimentador Artificial

El alimentador artificial es un dispositivo que sirve para proporcionar sangre o una solución alimenticia artificial al insecto de manera controlada (**Figura 1.2**). Consiste en un recipiente de alimentación (RA) que contiene la solución alimenticia, y un recipiente contenedor (RC) donde se aloja el insecto durante el proceso de alimentación. El RA es un recipiente de vidrio de 1 cm de diámetro y 2 cm de alto que posee en la parte inferior una membrana de látex, simulando la piel de un hospedador. El RC es un recipiente que se ubica debajo del RA, y que tiene en la parte superior una malla de tul para impedir la fuga del animal durante el ensayo. Esta tela permite que el insecto entre en contacto con la membrana de látex del RA, y acceda a la solución alimenticia tras perforarla mediante las mandíbulas de su rostro. La membrana se mantiene a *ca.* 34 °C simulando la piel de un hospedador.

1.2.4 Sistema de Calentamiento

Utilizando un baño termostatzado se recircula el agua mediante una bomba a través de mangueras de silicona conectando con el RA y calentando la solución contenida en su interior (**Figura 1.2**). El RA se calienta a una temperatura de 34 ± 2 °C, que es suficiente para provocar la extensión de probóscide de los insectos (Flores y Lazzari 1996).

1.2.5 Sistema de Registro de la Actividad Muscular

Un electrodo de registro se conecta a una malla metálica en el interior del RC sobre la cual se posa el insecto. El electrodo de referencia se sumerge en la solución alimenticia del RA.

Ambos electrodos convergen en un amplificador (EX1, Dagan Corporation) para incrementar la amplitud de la señal 200 veces (200x). Esta señal analógica es posteriormente digitalizada mediante un conversor analógico/digital (Tektronix, TDS 210) y enviada a una computadora para su visualización (**Figura 1.2**).

La actividad eléctrica de los músculos suctores de la cabeza del insecto se midió mediante el registro electromiográfico. La señal se produce cuando el animal cierra el circuito eléctrico al insertar los estiletes en la solución contenida en el RA (**Figura 1.2**). La adquisición de los registros individuales es posible gracias a un código desarrollado *ad hoc* en Matlab. Los registros generados permiten observar los cambios de voltaje asociados a la actividad de los músculos suctores y distinguir otros comportamientos característicos del proceso de alimentación asociados a un patrón reconocible en la señal (**Figura 1.3**).

1.2.6 Diseño Experimental

El experimento se iniciaba una vez colocado un insecto en el RC del alimentador artificial. Se le otorgó al insecto un tiempo máximo de 2 minutos para que entre en contacto con la solución alimenticia a través de la membrana. Si el insecto no se alimentaba en ese período, era retirado y reemplazado por otro. En caso de iniciarse la alimentación, el insecto permanecía en el recipiente durante 10 minutos, lo cual le permitía ingerir el alimento a repleción. Todos los experimentos se realizaron al inicio de la escotofase del insecto (0 - 6 h después del apagado de las luces), momento del día en el que los triatomíneos muestran su máxima motivación para buscar una fuente de alimento (Barrozo *et al.*, 2004; Bodin *et al.*, 2008).

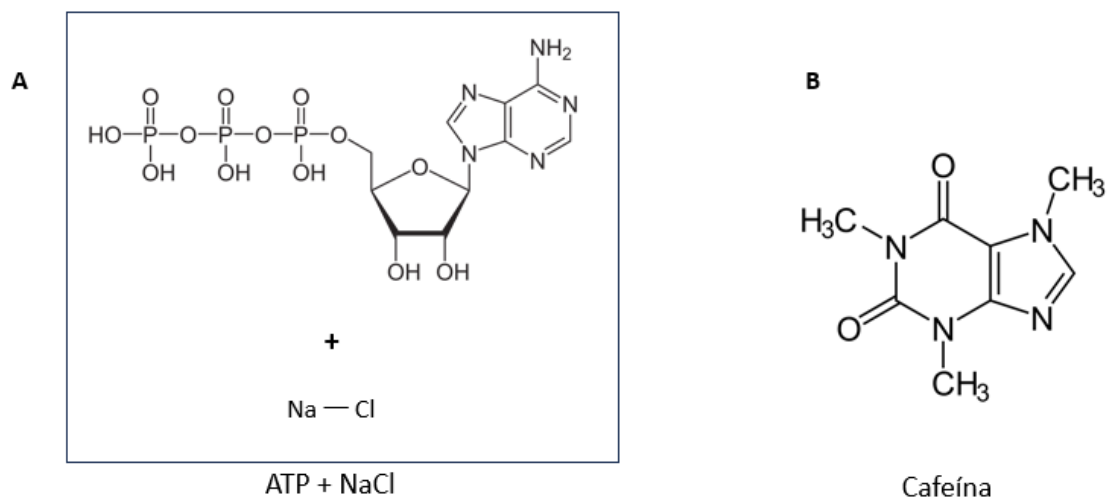


Figura 1.1. Compuestos utilizados en los experimentos. **A** - Compuesto apetitivo: ATP + NaCl . **B** - Compuesto aversivo: cafeína.

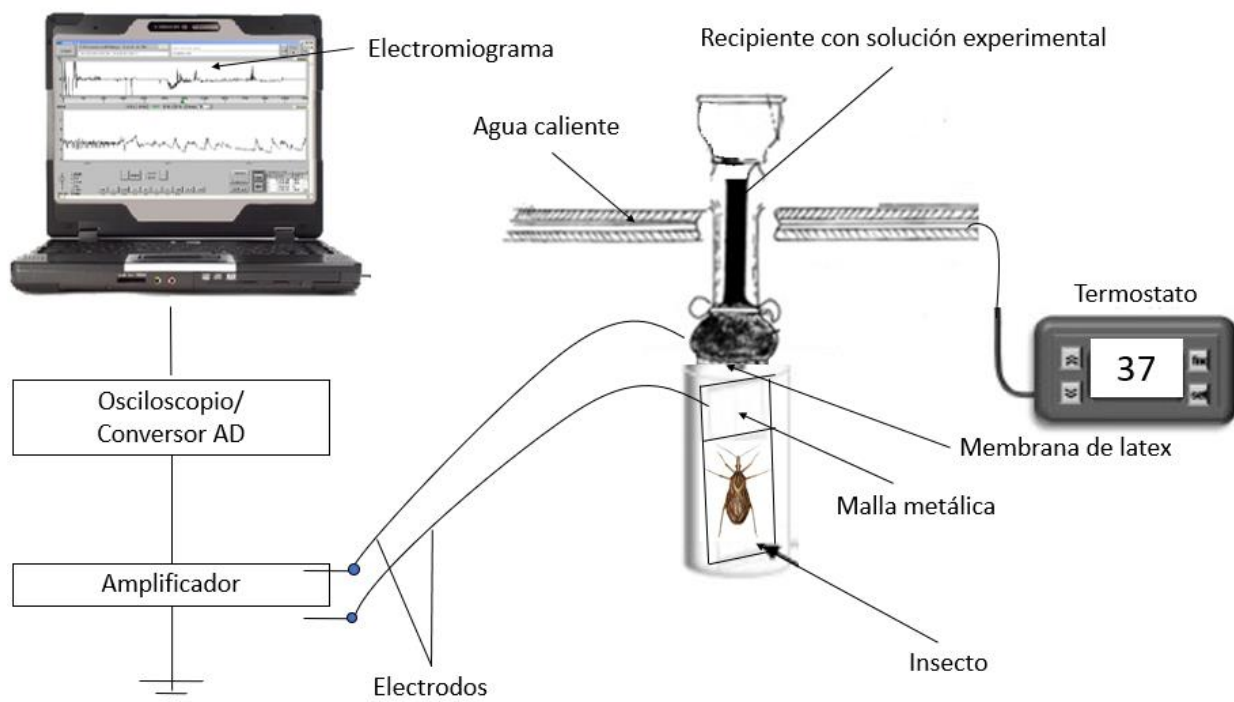


Figura 1.2. Sistema de registro de la actividad electromiográfica durante la alimentación. Alimentador artificial compuesto por un recipiente de alimentación (RA) con la solución a testear, y un recipiente contenedor (RC) del insecto. El RA tiene una membrana de látex en su extremo inferior, simulando la piel de un hospedador. El insecto perfora la membrana para acceder a la solución del RA. El RA y la membrana se mantienen a 34 °C con la ayuda de un baño

termostatzado. La malla metálica en el RC y el electrodo inmerso en el RA están conectados al amplificador. La señal de los electromiogramas (EMGs) es digitalizada y almacenada en una PC.

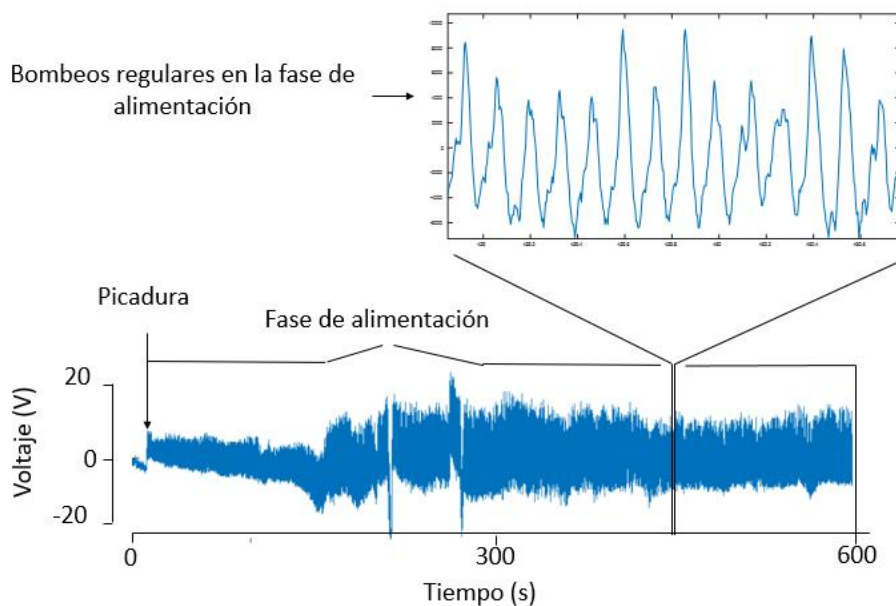


Figura 1.3. Electromiograma de *Rhodnius prolixus* durante la ingestión de una solución apetitiva (ATP en 0,15M NaCl).

1.2.7 Análisis Estadístico

Las variables se analizaron estadísticamente utilizando la prueba de Mann-Whitney, con un nivel de significancia del 5 % ($\alpha = 0,05$) para evaluar la hipótesis nula (H_0 : No existen diferencias significativas entre tratamientos).

1.3 Resultados

1.3.1 Caracterización del Comportamiento de Alimentación

El insecto localizado en el alimentador artificial y en presencia de una fuente de calor realiza una serie de comportamientos estereotipados. El insecto se aproxima a la membrana de látex del RA y extiende su probóscide formando un ángulo recto con su superficie, este comportamiento es disparado por el calor emanado por el RA y la motivación propia del insecto.

Si el insecto perfora la membrana del RA, los estiletes entran en contacto con la solución del RA. En la **Figura 1.3** se señala la “picadura” de la membrana, que se identifica por un cambio de voltaje en el registro del EMG. Este cambio conspicuo de voltaje corresponde al momento en que el circuito eléctrico se cierra, produciéndose una continuidad eléctrica entre la solución del alimentador, que contiene el electrodo de referencia, y el insecto, quien está en contacto con el electrodo de registro. El tiempo transcurrido desde que el insecto inserta sus estiletes en el RA hasta que los extrae del mismo lo denominamos **Período de Contacto (PC)**. Dentro del PC, usando sus estiletes y realizando pequeñas contracciones de las bombas cibarial y faríngea, el insecto absorbe pequeñas cantidades de líquido. Durante este tiempo, el insecto posiblemente evalúe su calidad sin que se produzca una verdadera ingestión, este período se denomina **Fase de Muestreo (FM)**. Si el alimento reúne las cualidades requeridas, el insecto comienza con la fase de alimentación mediante la succión continua y estable de la solución hacia el sistema digestivo (**Figura 1.3**). Esta fase se caracteriza por contracciones regulares y de amplitudes mayores que en la FM de los músculos suctores (**Figura 1.3**). A partir de este momento comienza la **Fase de Verdadera Ingestión (FVI)**. Este proceso se puede interrumpir, ya sea cuando el insecto succiona la máxima cantidad de alimento y su abdomen se expande a repleción, o parcialmente. En caso de que la alimentación sea parcial el insecto puede iniciar un nuevo período de evaluación y muestreo, de verdadera ingestión, o terminar el contacto con la solución mediante la extracción de los estiletes. Todos los eventos descritos previamente pueden repetirse varias veces durante el experimento, y son reconocibles en los EMGs debido a las particulares deflexiones de voltaje que se producen. No obstante, estos cambios de voltaje pueden presentar variaciones significativas en forma y duración, dependiendo del tipo de alimento disponible para el insecto.

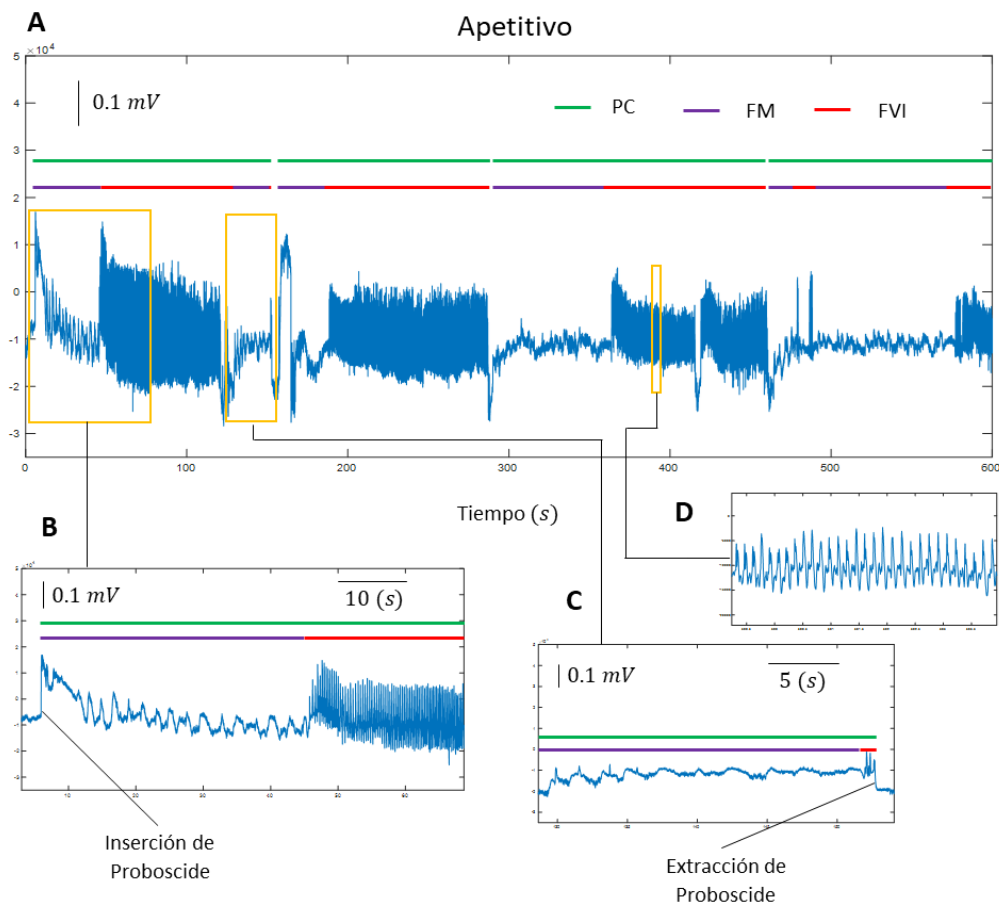
1.3.2 Comportamiento de Alimentación hacia una Solución Apetitiva y una Aversiva

El comportamiento alimenticio de los triatominos en el alimentador artificial revela características diferenciables cuando el animal ingiere una solución apetitiva o una aversiva. En la **Figura 1.4** se observan dos ejemplos de EMGs cuando un insecto se alimenta con cada una de estas soluciones.

La FM empieza con el despliegue de los estiletes en la solución del RA y la posterior ingesta de pequeñas cantidades de alimento. La succión genera picos de voltaje con formas y tamaños que pueden ser irregulares, pero que, en general, suelen presentar una amplitud que varía entre los 0,03 - 0,1 mV, y muestran un mayor ancho con respecto a los observados en la FVI. Los picos de voltaje que generalmente se exhiben durante la FVI son más angostos que durante la FM y de amplitudes mayores, que varían entre 0,1 - > 0,2 mV. La inspección de los registros revela que la FM no ocurre exclusivamente al inicio de la alimentación y puede repetirse sucesivas veces durante la misma. El insecto luego de una FM puede iniciar la FVI y hacer 1 o varias pausas, y, sin retirar los estiletes del alimentador comenzar una FM (**Figura 1.4A**). Es factible que durante estas FMs, que pueden variar en tiempo (desde 5 hasta 60 segundos) y número (desde 1 hasta más de 5) según el registro y solución ofrecida, el insecto realice evaluaciones gustativas.

Tanto la FM como la FVI están incluidas en el período de contacto (PC), que comprende el tiempo efectivo que los insectos tienen sus piezas bucales dentro del alimentador. En el ejemplo de la **Figura 1.4A**, se observa para la solución apetitiva, que el registro puede estar fragmentado de la siguiente manera: el insecto inserta sus estiletes en el alimentador y comienza el primer PC, este comprende 1 evento de muestreo (FM), 1 evento FVI seguido de 1 FM; el insecto retira sus estiletes y los vuelve a insertar en un segundo PC, que comprende 1 evento FM y 1 evento FVI; el insecto nuevamente sustrae y vuelve a insertar los estiletes y comienza el tercer PC que es similar al segundo PC; hacia el final del registro el insecto retira y reinserta las piezas bucales y da lugar al cuarto PC, este está comprendido por 2 FM y 2 FVI. En el ejemplo representativo del comportamiento de un insecto hacia una solución aversiva de la **Figura 1.4E**, se observa en cambio, una primer FM seguida del retiro y reinsertación de los estiletes que siguen a una segunda FM seguida de una FVI; el insecto luego repite la acción anterior 2 veces y en un quinto evento de inserción de las piezas bucales el insecto solamente realiza 1 FM para luego retirar definitivamente los estiletes. Es notable que cuando el insecto se alimenta de una solución aversiva, frecuentemente, los picos de voltaje pueden ser de menor amplitud que hacia la solución apetitiva (**Figura 1.4E-G**). Cuando al insecto se le ofrece una solución apetitiva, la duración de la FVI normalmente abarca al menos 50 % del tiempo que dura el experimento (**Figura 1.4A**). La FVI que exhiben los insectos hacia soluciones apetitivas suele incluir una

cantidad de picos o contracciones musculares que muchas veces superan las 500 unidades y en algunos casos se cuentan en más de 3000. En cambio, en los insectos expuestos a una solución aversiva la FVI es más corta (< 50% del tiempo del experimento), con un número de picos visiblemente menor y una duración de la motivación por alimentarse notablemente inferior que hacia la solución apetitiva.



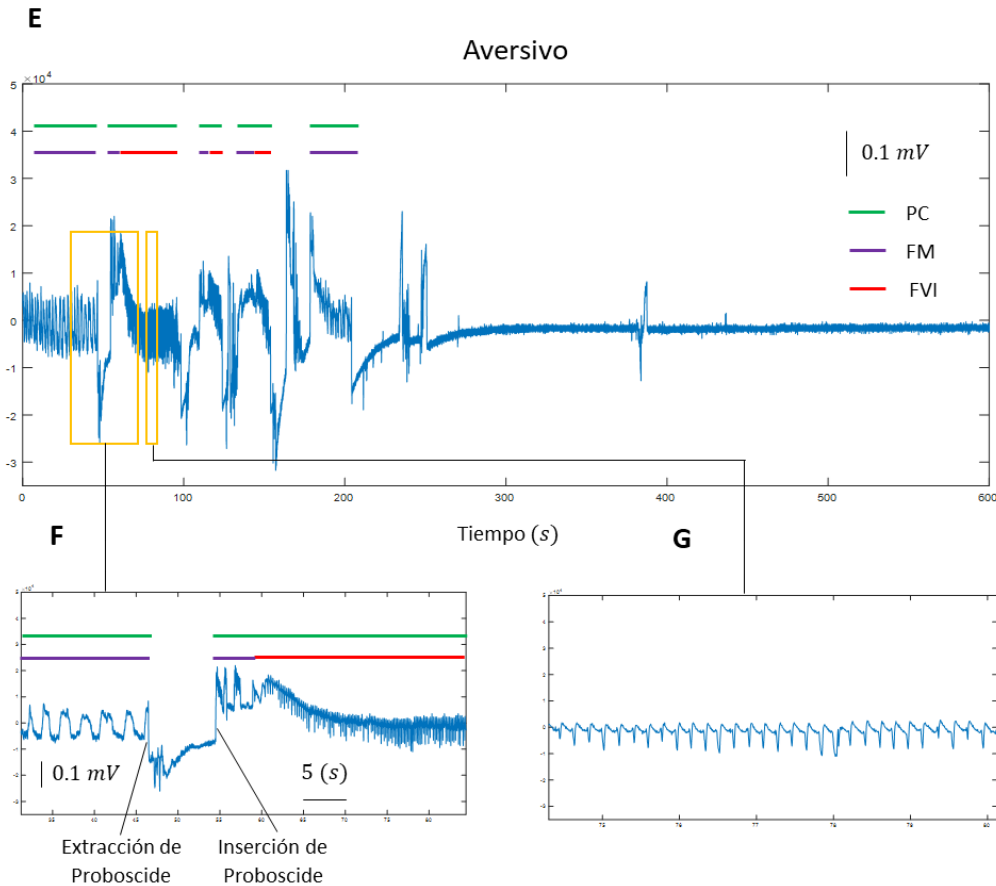


Figura 1.4. Electromiogramas producidos por una solución apetitiva y una aversiva. Las líneas verdes, púrpuras y rojas indican el instante de ocurrencia y la duración del Periodo de Contacto (PC), la Fase de Muestreo (FM), y la Fase de Verdadera Ingestión (FVI) respectivamente. **A** - EMG obtenido al alimentar una ninfa de *R. prolixus* con una solución apetitiva compuesta por ATP y NaCl. **B** - Segmento ampliado del EMG hacia la solución apetitiva donde se observa la inserción de los estiletes del insecto en la solución, así como deflexiones de voltaje típicas en la FM seguidas de la FVI. **C** - Segmento ampliado del EMG hacia la solución apetitiva donde se observan deflexiones de voltaje típicas de la FM, seguidas por una FVI muy breve y la posterior extracción de la probóscide. **D** - Segmento ampliado del EMG apetitivo donde se observan los picos de voltaje asociados al bombeo de la solución apetitiva durante la FVI. **E** - EMG obtenido hacia la solución aversiva compuesta por cafeína, ATP y NaCl. **F** - Segmento ampliado del EMG hacia la solución aversiva donde se observan los eventos de inserción y extracción de probóscide en la solución, así como deflexiones de voltaje ocurridas durante la FM y la FVI. **G** - Segmento ampliado del EMG hacia la solución aversiva donde se observan los picos de voltaje asociados al bombeo de la solución durante la FVI.

La **Figura 1.5** muestra la dinámica temporal del comportamiento de los insectos expuestos hacia una solución apetitiva y hacia una aversiva. En la **Figura 1.5A** se observa cómo cambia el número de insectos que efectivamente tienen sus estiletes dentro del alimentador, es decir, que se encuentran en PC independientemente si se alimentan o no y en la **Figura 1.5B**, se muestra la variación temporal del número de insectos que se encuentran en una FVI. Las diferencias son notables entre ambos grupos de insectos expuestos a las diferentes soluciones. El número de insectos que invierte tiempo alimentándose o simplemente en contacto con la solución aversiva disminuye más rápido que aquellos insectos que se alimentan de una solución apetitiva.

La sumatoria de todas las FM de cada registro nos brinda información de la duración total del muestreo (DM) para cada individuo. Asimismo, es posible cuantificar la duración total de las ingestas (DI) como la suma de todos los eventos de FVI para cada insecto. En la **Figura 1.6** se grafica el PC, la DM y la DI que muestran los insectos hacia la solución apetitiva y hacia la solución aversiva. El análisis comparativo de los resultados revela que el PC es significativamente mayor hacia la solución apetitiva con respecto a la aversiva (PC apetitivo vs. aversivo: Mann-Whitney = 130, $gl = 1$, $p = 0,0073$) y marginalmente significativo para la DI entre ambas condiciones (DI apetitivo vs. aversivo Mann-Whitney = 163, $gl = 1$, $p = 0,057$). Es decir, los insectos que se alimentaron de una solución apetitiva pasaron más tiempo en contacto con la solución (mayor PC) y en particular pasaron más tiempo ingiriendo (mayor DI) que aquellos que se les ofreció una solución aversiva.

De los resultados hasta aquí mostrados se desprende la factibilidad de reconocer diferentes patrones alimenticios hacia soluciones con diferente valencia. Además, remarcan el carácter antialimentario y de rechazo que produce en los insectos la solución aversiva observables en los EMGs.

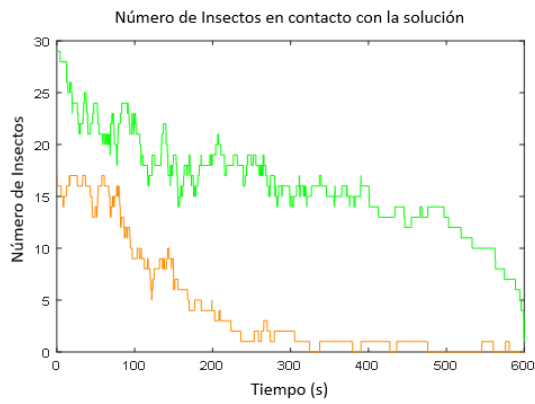
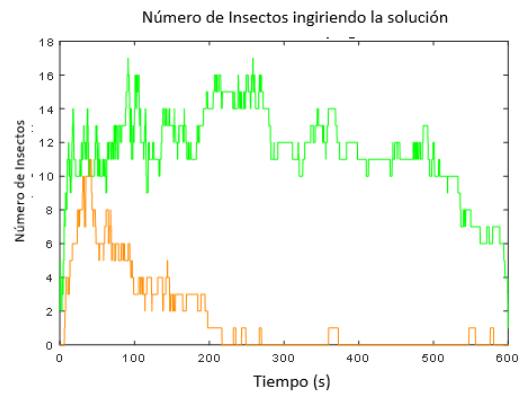
A**B**

Figura 1.5. Dinámica temporal del comportamiento alimenticio hacia una solución apetitiva (curva verde) y una aversiva (curva naranja). **A** - Número de insectos en contacto con la solución. **B** - Número de insectos ingiriendo la solución.

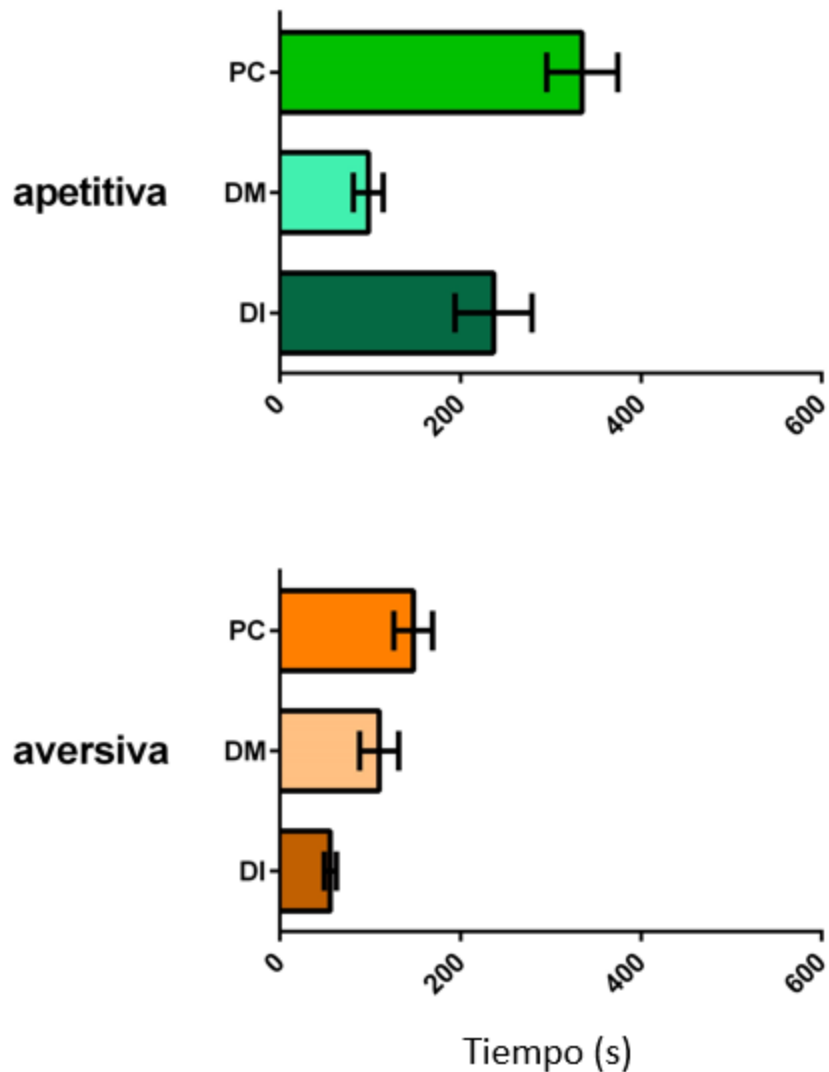


Figura 1.6. Duración de las fases que componen el comportamiento de alimentación de *R. prolixus*. PC, es el tiempo que el insecto mantiene sus estiletes en el alimentador, DM, es el tiempo que dura el muestreo, y DI, es la duración total de la ingestión efectiva. Los insectos fueron expuestos en el alimentador artificial hacia una solución apetitiva **(A)** y una solución aversiva **(B)**. Se representan los valores medios (+/- EE). El número de individuos evaluados para cada solución tratamiento fue 29 y 17, respectivamente.

1.4 Discusión

En este capítulo identificamos y describimos el comportamiento alimenticio de *R. prolixus* hacia dos tipos de soluciones, por un lado, una apetitiva y por otro una aversiva, basándonos en la inspección y análisis de los EMGs. Al observar los EMGs obtenidos para cada tratamiento, es posible reconocer cualitativamente efectos distinguibles que provocan comportamientos diferentes a la escala de los músculos suctores. Cuando el alimento es apetitivo o fagoestimulante, el insecto permanece mayormente inmóvil en el alimentador artificial y realiza inserciones y extracciones de los estiletes con una frecuencia menor respecto a los casos donde el alimento es aversivo. Esto hace que el EMG sea menos ruidoso y los cambios de voltaje se deban a la actividad de las bombas cibarial y faríngea de forma predominante. Con una solución aversiva el insecto abandona la alimentación en la primera mitad de la duración del experimento, lo que implica que las inserciones, aunque igualmente numerosas, se producen durante ese período restringido. Además, observamos que el PC es dos veces superior hacia la solución apetitiva y que la DM, a pesar de tener la misma duración absoluta con ambos tratamientos, respecto al PC tiene una proporción mayor para la solución aversiva. En otras palabras, el insecto utiliza la mayor parte del PC probando el alimento cuando se le ofrece la solución aversiva, a diferencia de lo que ocurre con la solución apetitiva, donde el PC está conformado principalmente por la fase de verdadera ingestión (FVI). En consecuencia, los eventos de FVI tienen una duración absoluta y relativa considerablemente menor con la presencia de la solución aversiva, lo que se interpreta como una inhibición del proceso de alimentación.

Durante la fase de muestreo (FM), con ambos tratamientos, se producen picos de voltaje de tamaño similar. Es notable encontrar una frecuencia de aproximadamente 0,5 Hz que coincide con las emisiones de saliva descritas previamente para esta fase cuando el insecto se alimenta de un hospedador (Soares *et al.*, 2006). Durante la FVI, es posible observar picos dobles que corresponden a movimientos de contracción y relajación periódicos y asíncronos entre los músculos que conforman la bomba cibarial y la bomba faríngea del insecto, mientras se alimenta. Sin embargo, al igual que lo reportado por Smith (1979), la aparición de dos o incluso más picos solapados con formas diferentes en ocasiones parece no tener un patrón aparente. Esto puede

deberse a la presencia de contracciones débiles, interrumpidas o irregulares en las bombas, que se producen a pesar de que la succión del alimento durante las FVIs es notablemente más estable y sostenida que en las FMs (**Figura 1.4G**). En consecuencia, los picos de menor tamaño no son siempre visibles. Con algunos tratamientos, los cambios en la conductividad del alimento son más sensibles a la apertura de válvulas en el canal salival, lo que permite que los picos de voltaje asociados a estos eventos sean más notables y sean visibles durante la fase de verdadera ingestión. Este efecto ha sido reportado en insectos áfidos (McLean y Kinsey, 1967).

En este capítulo se describió el comportamiento de *R. prolixus* hacia diferentes soluciones de manera cualitativa mediante la visualización de los EMGs. Esto representa un paso crucial para determinar cuáles son las fases de alimentación y el comportamiento esperable para situaciones alimenticias opuestas, con el fin de poder establecer los parámetros necesarios en el desarrollo de un programa que permita la identificación de los distintos comportamientos de manera automática o semiautomática. La identificación manual de patrones de onda, comportamientos, fases de alimentación en los EGMs, puede tomar mucho tiempo para completarse y aumentar la incidencia de errores humanos, lo que produce un "cuello de botella" en el proceso de análisis (Prufer *et al.*, 2014). Es decir, al analizar los EMGs de la bomba cibarial y faríngea de los insectos triatominos, es necesario marcar los límites donde ocurren los comportamientos de interés, como por ejemplo, el PC (delimitado por las inserciones y extracciones de los estiletes), la FM y la FVI. Una vez que se marcan los límites de estos eventos, se necesita contar los picos de voltaje para calcular la frecuencia de bombeo. Algunos registros EMGs pueden contener cientos e incluso miles de picos a contabilizar, sumado a las réplicas necesarias para realizar comparaciones estadísticas entre tratamientos, donde se suelen realizar alrededor de 20 registros electromiográficos o más por cada uno. Esto implica analizar al menos 40 EMGs para comparar estadísticamente 2 tratamientos. Adicionalmente, medir las amplitudes de los picos sería una tarea aún más laboriosa debido a que la señal comúnmente tiene una línea basal irregular, por lo que no se puede utilizar una referencia de voltaje fija, siendo imprescindible la obtención de los valores de potencial de la parte superior e inferior de cada uno de los picos para calcular su diferencia, resultando en 3 operaciones por cada uno (medición de la altura, el valle y el cálculo de la diferencia). Por lo tanto, la necesidad de utilizar una herramienta computacional

para que el proceso sea más eficiente y rápido es evidente. En el siguiente Capítulo se abordará el desarrollo de este programa computacional para agilizar el análisis de la información extraída de los EGMs.

Capítulo 2. MyoRec: Programa Computacional para Optimizar el Estudio del Comportamiento de Alimentación

PARTE A: Identificación de Patrones de Comportamiento Mediante Modelos de Machine Learning y Deep Learning

2.A.1 Introducción

Una de las grandes limitaciones de la técnica de electromiografía en insectos, y especialmente en triatominos, es la ausencia de una herramienta computacional de fácil acceso que permita disminuir el tiempo de análisis de los EMGs y facilite el manejo de las grandes cantidades de datos que pueden surgir cuando se procesa un número elevado de estos registros. Aunque las aplicaciones desarrolladas hasta ahora se pueden usar para extraer y procesar datos de EMGs de diferentes insectos, en la mayoría de los casos fueron desarrolladas específicamente para áfidos y tienen algunas limitaciones cuando se utilizan en animales diferentes. Estas limitaciones se derivan principalmente de la variabilidad de los patrones de onda que componen las fases de alimentación. Varias de estas aplicaciones cuentan con un sistema de reconocimiento automático de algunos comportamientos pero su efectividad puede variar significativamente cuando se usa en insectos diferentes, incluso en especies diferentes de áfidos (Adasme-Carreño *et al.*, 2015). Esto se traduce en una pérdida de las ventajas que ofrecen estos mecanismos y en un aumento del tiempo de análisis. Por otro lado, la visualización de cambios dinámicos en los parámetros durante el tiempo de registro puede proporcionar información útil sobre el proceso de alimentación, pero los datos son difíciles de almacenar y organizar por su volumen masivo y las variaciones en el número y duración de los eventos asociados al comportamiento alimenticio. A pesar de esto, no existe un programa especializado que permita gestionar, almacenar y visualizar este tipo de datos de forma eficiente.

El objetivo de este Capítulo es desarrollar un sistema de reconocimiento automático de patrones para la caracterización del comportamiento de alimentación (*i.e.*, fases de verdadera ingestión, FVIs) en insectos triatominos, que estará contenido en un programa o *software*, que denominamos MyoRec, destinado a optimizar y facilitar la extracción, el almacenamiento, el procesamiento, la visualización y la comparación de datos de EMGs. El *software* constará de tres módulos principales: 1) Identificador de Eventos de Verdadera Ingestión (IEVI); 2) Detector de Picos (DP) y 3) Calculador de Datos (CD).

En la **Parte A** del presente Capítulo nos centraremos únicamente en el desarrollo del IEVI, abarcando los módulos restantes y otros detalles del programa en la **Parte B** del capítulo. El desarrollo del IEVI responde a la necesidad de automatizar tanto como sea posible el proceso de análisis de los EMGs, y ayudar al investigador a reducir el tiempo que utiliza al delimitar los comportamientos relevantes, en particular los eventos FVI. Nótese que un investigador avezado puede tardar de 5 a 20 minutos por registro (a veces incluso más tiempo) en identificar y establecer manualmente todos los eventos asociados a comportamientos de alimentación, dependiendo de la complejidad del registro.

2.A.1.1 Antecedentes de Programas Computacionales Usados en el Análisis de Electromiogramas en Distintos Insectos

Los primeros programas computacionales se orientaron a la automatización de los análisis de datos provenientes de EPGs. Las investigaciones comenzaron a finales de la década de los 80 y se enfocaron principalmente en insectos fitófagos como los áfidos, que representan uno de los grupos de plagas de mayor importancia económica en la agricultura (Guerrieri y Digilio, 2008). En este sentido, Tjallingii (1988) realizó los primeros intentos de reconocimiento automático de patrones de EPGs utilizando el análisis de Fourier, pero fueron insatisfactorios debido a la variabilidad biológica de la señal (Rahbé *et al.*, 1995). A principios de los 90 se desarrolló el primer programa *ad hoc*, Stylet 1.0, que facilitaba la adquisición y el análisis de datos. Evolucionó a lo largo de los años con otros programas relacionados aumentando la velocidad de procesamiento, ampliando el número de variables calculadas e integrando herramientas de análisis estadístico eficientes (Van Giessen y Jackson, 1998, Sarria *et al.*, 2009, Giordanengo, 2014, Adasme-Carreño *et al.*, 2015, Elbert *et al.*, 2015). También se produjeron avances notables en el reconocimiento

de patrones en tiempo real para identificar las fases de alimentación (Kasetty *et al.*, 2008). Desarrollos posteriores como A2EPG (Adasme-Carreño *et al.*, 2015), Aphid-AutoEPG (Prufer *et al.*, 2014) y los algoritmos de aprendizaje automático (Willett *et al.*, 2014) mejoraron el reconocimiento automático de patrones de onda, aunque sin habilitar aún capacidades para el procesamiento total de los análisis en tiempo real.

En el caso de los triatominos, se han desarrollado herramientas para la estimación automática de variables (Araujo *et al.*, 2011), pero no existe en la actualidad un *software* accesible para automatizar el reconocimiento de comportamientos y el cálculo automático de variables.

2.A.1.2 Modelos Matemáticos para el Reconocimiento de Patrones en el Identificador de Eventos de Verdadera Ingestión

Un sinnúmero de modelos matemáticos han demostrado ser efectivos en el reconocimiento de patrones cuando se aplican a diversos tipos de datos, como imágenes, texto o series temporales. Para obtener un adecuado Identificador de Eventos de Ingestión (IEVI) que identifique y delimite automáticamente las fases de alimentación en los EGMs, debe aplicarse uno o varios modelos mediante algoritmos adecuados. Para cumplir con tal propósito se compararon los siguientes tipos de modelos: regresiones lineales y logísticas, árboles de decisión, ensambles (*Voting*, *Random Forest* y *Gradient Boosting*), redes neuronales (que incluyen redes convolucionales y recurrentes) y Máquinas de Soporte Vectorial. En el Apéndice A se detallan y describen los fundamentos generales de cada uno de ellos, enmarcados en el ámbito de *machine learning* y de *deep learning*. Algunos modelos son bastante efectivos por sí solos, pero es posible incrementar su desempeño escogiendo los parámetros adecuados, determinadas arquitecturas y al generar modelos híbridos (stacks y ensambles) que compensen sus debilidades individuales.

2.A.2 Materiales y Métodos

2.A.2.1 Datos de Entrenamiento y Validación

Para desarrollar el Identificador de Eventos de Verdadera Ingestión (IEVI), se generaron dos conjuntos de datos para el entrenamiento de los diversos modelos de *machine learning* y *deep learning*. Estos conjuntos de datos se crearon a partir de la división secuencial de los EMGs en ventanas temporales de un segundo de duración. Cada ventana (instancia) consistió en 250 muestras (medidas puntuales que conforman la señal) con una frecuencia de muestreo de 250 Hz (**Figura 2.1**). El primer conjunto de datos, denominado *Raw Data*, está representado por los segmentos del EMG delimitados por las ventanas. El segundo conjunto de datos, llamado *FFT Data*, es similar al primero, con la diferencia que se aplica la transformada de Fourier a cada una de las instancias que tienen, esta vez, 126 muestras.



Figura 2.1. Estructura de los datos *Raw* y *FFT* utilizados para el entrenamiento y validación de los modelos candidatos del Identificador de Eventos de Verdadera Ingestión (IEVI). Las clases de instancias (Con picos y Sin picos) en los conjuntos de datos *Raw* y *FFT* se establecen de forma aleatoria y se normalizan. Ambos conjuntos de datos se balancean para que tengan el mismo número de instancias de ambas clases antes de que se suministren a un modelo.

Al realizar el entrenamiento con los datos, el algoritmo del IEVI identifica dos clases: ventana con picos de voltaje, y ventana sin picos. Los intervalos temporales que constituyen los eventos de la fase de verdadera ingestión (FVIs), se demarcaron manualmente mediante MyoRec. Una vez establecidos los límites de dichos eventos, se clasificaron las instancias dentro y fuera de ellos con la etiqueta correspondiente “Con Picos” y “Sin Picos”. Las instancias que llegan a contener a las coordenadas de estos límites fueron clasificadas como “Sin Picos” cuando más de la mitad de su longitud se encuentra por fuera del evento, y de forma contraria (“Con Picos”) si se encuentra dentro de los límites del evento. Finalmente los datos de las instancias se normalizan y se introducen en el modelo de forma aleatoria, el cual debe aprender a reconocer ambas clases de ventanas durante el proceso de entrenamiento.

Para disminuir la probabilidad de que el reconocimiento de las clases por parte del modelo estuviese sesgada, se balanceó el número de instancias pertenecientes a cada una para los conjuntos de datos (*Raw* y/o *FFT*) que serían usados en su entrenamiento. La técnica de balanceo utilizada se conoce como *undersampling*, y consiste en reducir el número de instancias de la clase mayoritaria a la clase minoritaria. Los conjuntos de datos *Raw* y *FFT* utilizados contaban, antes de ser procesados, con un número mayor de instancias pertenecientes a la clase “Sin Picos”. El resultado final fue que los conjuntos de datos *Raw* y *FFT* se constituyeron equitativamente con un número similar de instancias de ambas clases.

Para facilitar la identificación de patrones en los EGMs y acelerar el aprendizaje en el entrenamiento del modelo, se utilizó el algoritmo *Base Line Estimation and Denoising With Sparsity* (BEADS), con el objetivo de eliminar la tendencia y las irregularidades de la línea basal de la señal (Ning *et al.*, 2014) (**Figura 2.2**). Un total de 60 EGMs generados tras alimentar a los insectos con compuestos apetitivos y aversivos, se utilizaron para generar las instancias, lo que produjo un total de 27376 ventanas temporales en cada conjunto de datos (*Raw* y *FFT*). Posteriormente éstos fueron divididos en una sección con un 80 % de los datos para el entrenamiento y validación de los modelos (21900) y otra con el 20 % restante para el proceso de testing (11379).

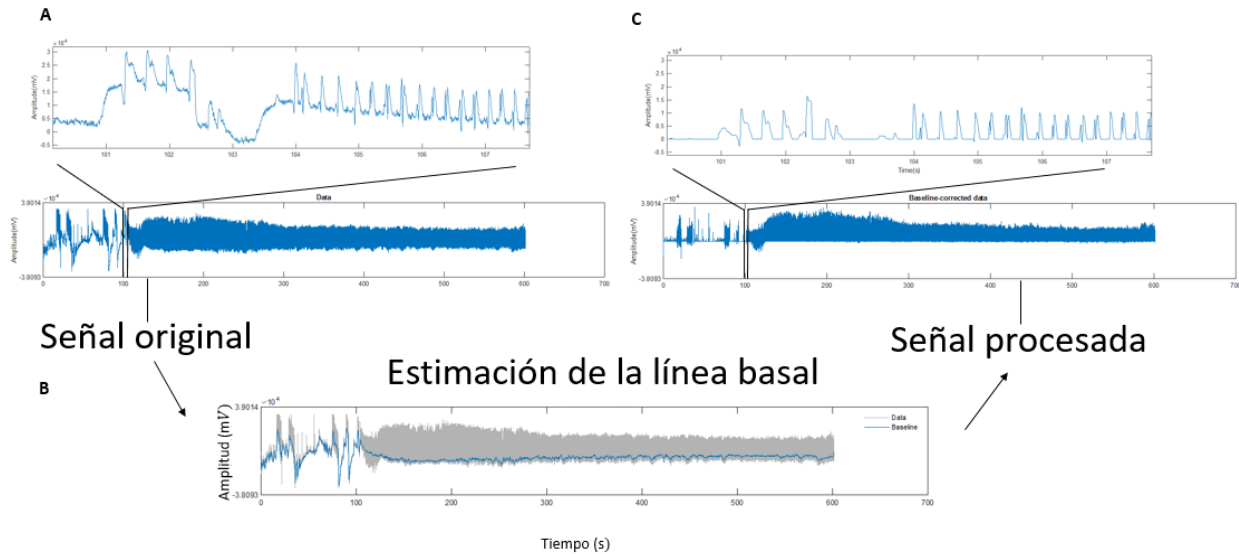


Figura 2.2. Aplicación del algoritmo *Base Line Estimation and Denoising With Sparsity* (BEADS) a un EMG. **A** - Señal original con foco en uno de sus segmentos. **B** - Estimación de la línea basal en la señal original (azul) mediante la aplicación de BEADS. **C** - Señal resultante al aplicar el algoritmo BEADS a la señal original. La línea basal estimada en **B** se sustrae de la señal original, produciendo una línea basal más regular y aproximadamente constante en el valor de voltaje cero.

2.A.2.2 Modelos de Clasificación Probados para el Identificador de Eventos de Verdadera Ingestión

Una serie de modelos de clasificación fueron entrenados con los conjuntos de datos previamente descritos para determinar cuál ofrece el mejor desempeño. Entre ellos destacan modelos de regresión logística, árboles de decisión, ensambles (que incluyen los tipos: votación por mayoría, *Random Forest* y *Gradient Boosting*), redes neuronales (que incluyen redes convolucionales, recurrentes y redes de memoria a corto y largo plazo) y Máquinas de Soporte Vectorial. También fueron evaluados modelos híbridos, conformados por dos o más tipos de modelos, utilizando métodos sistemáticos (*Tree-structured Parzen Estimators*, Búsqueda Cuadrícula y Aleatorizada) y manuales para la selección de hiperparámetros*. Para comparar los modelos más sobresalientes se aplicaron 25 iteraciones de validación cruzada con tres pliegues sobre el conjunto de datos de entrenamiento (14600 muestras para entrenamiento y 7300 para validación).

*Hiperparámetros: son parámetros de entrada del modelo que fija el usuario

En el caso de las redes neuronales, se comprobó el desempeño de diferentes arquitecturas y configuraciones. Se diseñó una CNN de tipo I (CNNt1) con una arquitectura simple o con un número pequeño de capas, y una CNN de tipo II (CNNt2), con un mayor número y variedad de capas. La CNN de tipo II, a diferencia del resto de los modelos, admite simultáneamente los datos *Raw* y *FFT* y se evaluó tanto de forma individual, como en combinación con otros modelos. En particular, se evaluaron las combinaciones CNNt2-RF, CNNt2-GB, CNNt2-SVM, CNNt2-LSTM, donde se usó la red neuronal para extraer características relevantes de los datos, que luego serían suministradas como entradas del modelo siguiente.

En el **Apéndice A** se muestra información sobre los fundamentos de los modelos probados, mientras que en el **Apéndice C** se encuentran detalles sobre los hiperparámetros* de cada uno de ellos, la arquitectura usada para las redes neuronales y el código utilizado para obtener los resultados mostrados en la siguiente sección. .

2.A.2.3 Métricas

Para medir el desempeño de los modelos durante el entrenamiento con los diferentes conjuntos de datos, se utilizó la **exactitud** (del inglés, **accuracy**), considerando que estos estaban balanceados. El **Accuracy** representa, en el marco de nuestro problema, la cantidad de predicciones correctas al clasificar las ventanas “Con Picos”.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

donde TP (del inglés, *True Positives*) son las ventanas “Con Picos” clasificadas como tal por el modelo, mientras que los TN (del inglés, *True Negatives*) representan las ventanas “Sin Picos” clasificadas de la misma forma por el algoritmo. FP (del inglés, *False Positives*) son las ventanas “Sin Picos” clasificadas erróneamente como “Con Picos”, mientras que FN (del inglés, *False Negatives*) representa las ventanas clasificadas como “Sin Picos” cuando en realidad poseen picos.

2.A.3 Resultados

Se comparó el desempeño de varios tipos de modelos con el fin de incorporar el más adecuado en el Identificador de Eventos de Verdadera Ingestión (IEVI). Para identificar un posible efecto relacionado al preprocesamiento de los datos de entrada, cada modelo se ejecutó con los datos *Raw* y *FFT* por separado, con la excepción de la CNNt2 (**Figura 2.3**). En el proceso se utilizó la mejor combinación de hiperparámetros encontrada para cada caso. La **Figura 2.3** muestra que los modelos entrenados con datos *FFT* tuvieron un mejor desempeño en todos los casos, exceptuando a la red neuronal convolucional de tipo I. De estos resultados se desprende que el desempeño o **Accuracy** está influenciado por el pre-procesamiento del conjunto de datos utilizado.

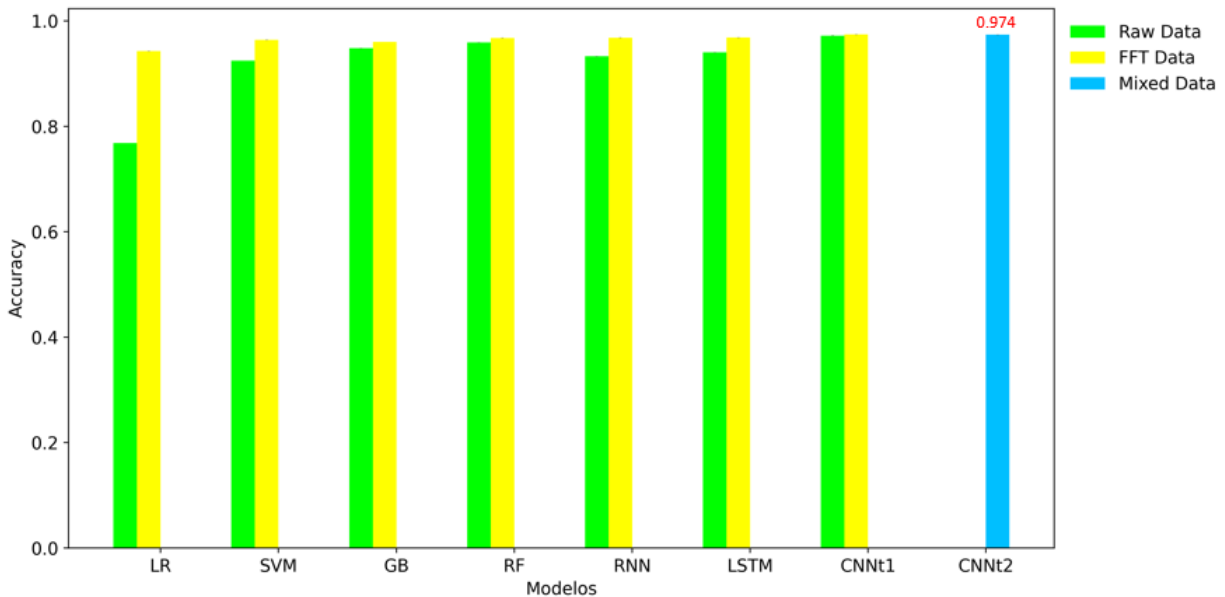


Figura 2.3. Accuracy para los diferentes modelos probados con los conjuntos de datos *Raw* y *FFT*. El modelo CNNt2 recibe como entrada ambos tipos de datos. El desempeño más alto se obtuvo con los dos tipos de CNNs (97.4 %).

Los modelos con mejor desempeño (**Accuracy** $\pm$ Error Estándar) fueron las redes neuronales convolucionales de ambos tipos (CNNt1 y CNNt2). Para la CNNt1 se obtuvo (97.2 % $\pm$ 0.02 %, 97.4 % $\pm$ 0.02 %) con los datos *Raw* y *FFT* respectivamente, y la CNNt2 registró (97.4 % $\pm$ 0.03 %). Para las redes de tipo LSTM se observaron *accuracies* de (94.0 % $\pm$ 0.02 %, 96.8 % $\pm$

0.02 %) con datos *Raw* y *FFT*, mientras que para las redes recurrentes (RNNs) se obtuvo ($93.3 \% \pm 0.04 \%$, $96.8 \% \pm 0.02 \%$). Con el ensamble de *Random Forest* los resultados fueron ($95.9 \% \pm 0.01 \%$, $96.8 \% \pm 0.01 \%$), y con *Gradient Boosting* ($94.8 \% \pm 0.01 \%$, $96.0 \% \pm 0.01 \%$). Finalmente, con los modelos SVM se obtuvieron los valores ($92.5 \% \pm 0.01 \%$, $96.4 \% \pm 0.009 \%$), mientras que con las regresiones logísticas se obtuvo ($76.8 \% \pm 0.02 \%$, $94.3 \% \pm 0.01 \%$).

De los resultados anteriores fueron escogidos los dos mejores modelos como candidatos para ser utilizados en el IEVI, es decir, la CNNt1 con datos de entrada *FFT* y la CNNt2 (Accuracy = 97.4 % en ambos casos). La complejidad, versatilidad y tiempo de entrenamiento son factores determinantes para escoger el modelo ganador. La CNNt1 tiene una arquitectura más simple que la CNNt2, y requiere datos de un solo tipo (*Raw* o *FFT*) como entrada. Ambos modelos tienen la capacidad de adaptarse a nuevos datos y aprender nuevas características sin tener que entrenarse desde cero. Sin embargo, el *accuracy* obtenido con la CNNt1 fue posible sólo mediante el uso de un valor bajo para el batch (uno de los hiperparámetros), lo que hace que el tiempo de entrenamiento sea mucho más largo que para la CNNt2. En la **Figura 2.4** se muestra el tiempo de entrenamiento promedio para ambos modelos al ser entrenados en 40 ocasiones. El tiempo de entrenamiento para el modelo CNNt1 fue de 1.78 ± 0.01 minutos, mientras que el del modelo CNNt2 fue de 0.30 ± 0.01 minutos. Esto significa que el procesamiento de ambos tipos de datos por parte del modelo CNNt2, no incrementa su tiempo de entrenamiento por encima del modelo CNNt1. Debido a que la obtención de los datos *FFT* a partir de los *Raw* es instantánea, tampoco producirá un incremento importante en la duración del tiempo de inferencia, es decir, el tiempo que el modelo necesita para identificar las muestras de un EMG una vez entrenado.

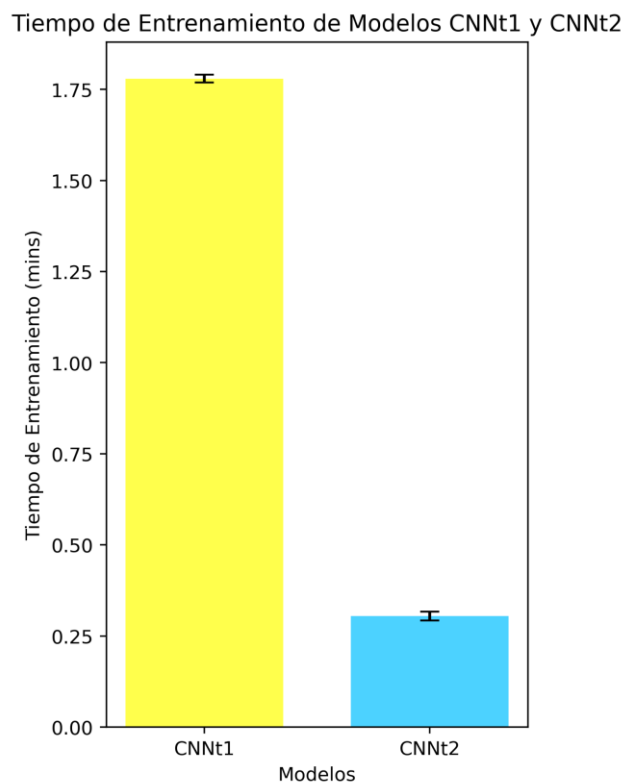


Figura 2.4. Tiempo de entrenamiento de los modelos CNNt1 y CNNt2 en minutos.

Los resultados preliminares para el **Accuracy** de los modelos combinados (CNNt2-SVM, CNNt2-RF, CNNt2-GB, CNNt2-LSTM, entre otros) fueron inferiores a los obtenidos con las CNNs por sí solas. Estos modelos también requerían mayor tiempo de entrenamiento, por lo que fueron excluidos del análisis.

En la **Figura 2.5** se muestra el diseño utilizado por la red CNNt2, seleccionada como el mejor modelo. Se utilizan dos conjuntos de datos como entradas (*Raw* y *FFT*) que son procesados a través de una secuencia de capas de convolución, normalización de instancias, *ReLU*, *MaxPooling* y *Dropout* con hiperparámetros que se adecúan al tamaño de los datos de entrada (250 muestras para los *Raw data*, y 126 para los *FFT*). Por tanto, los tipos y el número de capas que forman el diseño del modelo son similares para ambos conjuntos de datos, a diferencia de algunos hiperparámetros. Cada sucesión de capas consta de tres capas unidimensionales de convolución y dos de *MaxPooling*.

Luego del proceso de extracción de las características realizado en cada secuencia de capas de forma independiente, se concatenan y se introducen en otra serie de capas *Fully Connected* (FC) con funciones de activación *ReLus*. En cada una se utiliza un perceptrón multicapa* que modifica sus pesos para aprender los patrones identificados en las capas anteriores y posibilitar su posterior clasificación.

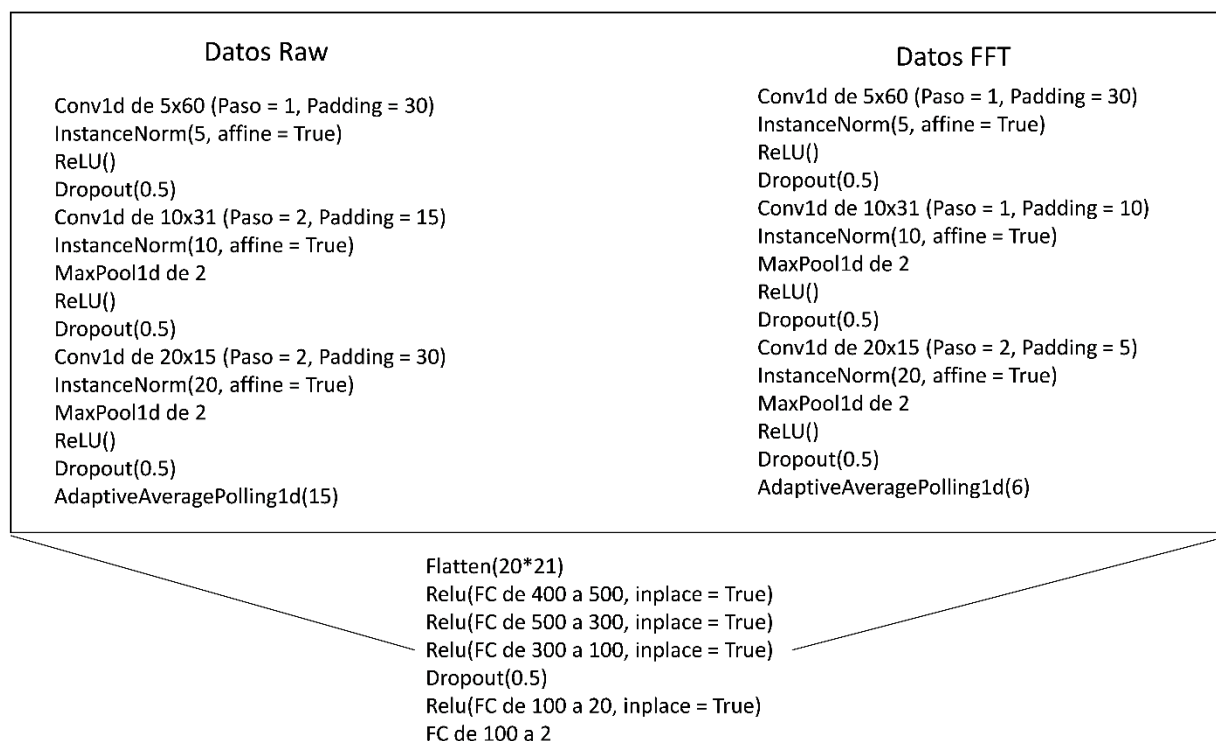


Figura 2.5. Modelo de Red Neuronal Convolutiva de tipo II que recibe segmentos de señales como entrada y clasifica dos tipos de patrones: con picos y sin picos. Este modelo fue incorporado en el IEVI debido a que tuvo el mejor desempeño.

*Perceptrón multicapa: es una red neuronal formada por más de dos capas.

2.A.3.1 Algoritmo de Refinación

Con el fin de mejorar el modelo y reducir el riesgo de falsos positivos y falsos negativos en el proceso de delimitación de las Fases de Verdadera Ingestión (FVIs), se elaboró un algoritmo de refinación. Esto mejora la detección de los límites de los eventos de verdadera ingestión (FVIs) y evita que se identifiquen más o menos eventos de los que existen realmente.

Una vez que se carga un archivo de datos *Raw* y se activa el IEVI, se aplica el algoritmo BEADS a la señal representada por el EMG y se elimina su media. Seguidamente, se realiza el proceso de clasificación donde se realizan tres iteraciones, asumiendo que los datos de entrada tienen longitudes diferentes en cada iteración. En la primera, se asume que los datos de entrada tienen longitudes de 750 y 376 muestras para los datos *Raw* y *FFT* respectivamente. En la segunda, se asumen los valores de 125 para los datos *Raw* y 63 para los *FFT*, mientras que en la tercera son de 64 y 32 respectivamente. Esto facilita al IEVI la identificación y delimitación de los eventos de verdadera ingestión con larga (>10 s), media (3 - 10 s) y corta duración (< 3 s). La aplicación de este algoritmo permite mejorar o refinar los límites de las FVIs. Los límites de las FVI identificadas sólo se modifican con las iteraciones siguientes si en ellas se establecen nuevos límites que engloban a los primeros. Para que un segmento de la señal se considere como un FVI, se estableció que debe conformarse de al menos tres ventanas con picos. Aún así, algunas FVIs con una duración de este orden pueden ser falsos positivos, especialmente en registros con baja relación señal/ruido. Para superar este problema, utilizamos el concepto de potencia de la señal aplicado a FVIs con una duración superior a cuatro segundos (1000 muestras) y calculamos su potencia promedio como valor umbral. En procesamiento de señales, se define la potencia de una señal discreta como la suma de los cuadrados absolutos de sus muestras en el dominio del tiempo dividida por la duración de la señal:

$$\overline{P_r(t)} = \frac{1}{t_2 - t_1} \sum_{r=t_1}^{t_2} |x(t)|^2$$

Donde $\overline{P_r(t)}$ es el umbral de potencia (o potencia promedio) del segmento de la señal (EMG) con límites t_2 y t_1 y $x(t)$ son sus muestras. $\overline{P_r(t)}$ se calcula promediando la potencia de todas las FVIs con duraciones mayores a 4 s. Las FVIs con una duración menor a 4 s y con una potencia por debajo de $\overline{P_r(t)}$, se consideran falsos positivos y por tanto se eliminan. El umbral de potencia ponderada también se utiliza para corregir los límites de las FVIs identificadas después del proceso de clasificación de ventanas. La potencia ponderada se compara con la potencia obtenida para un segmento (ventana temporal) de la FVI a partir del límite en cuestión. Si la potencia del segmento es mayor a la potencia promedio, el límite cambia de posición según la

longitud del segmento incrementando la duración de la FVI. Este proceso se repite de forma iterativa con segmentos externos a la FVI hasta que la potencia de uno de ellos cae por debajo del promedio. Para ajustar los límites reduciendo la duración de la FVI, se realiza un proceso similar pero con segmentos internos y con la regla inversa para la potencia, es decir, cambiando la posición del límite iterativamente hasta que la potencia del segmento sea mayor al promedio. El corrimiento máximo de los límites durante el ajuste es de 750 muestras.

Posteriormente se aplica un segundo proceso iterativo similar en el que se determina el valor más grande de la señal en los segmentos. Estos valores se comparan con un valor de referencia obtenido al calcular la amplitud promedio de los picos de voltaje en otro segmento, el cual tiene una duración equivalente al 20 % de la FVI y se extiende a partir del límite en cuestión. Durante el proceso, la posición de un límite se acerca a su homólogo cuando la altura promedio de los picos ponderada por un factor ajustable es mayor que el punto más alto de la señal en el segmento. El proceso termina cuando cambia la condición. Un límite se separa del otro si el punto más alto del segmento está por encima de la altura promedio ponderada de los picos detectados. Separar los límites es necesario, por ejemplo, cuando la duración de la FVI identificada es más reducida que su duración real. Todos los valores mencionados en esta sección pueden ser ajustados por el usuario.

El modelo CNNt2 se probó en numerosos EMGs apetitivos y aversivos. Su desempeño es satisfactorio en la mayoría de los registros apetitivos (**Figura 2.6**), pero aún es posible realizar ajustes adicionales para mejorar su efectividad en los registros aversivos.

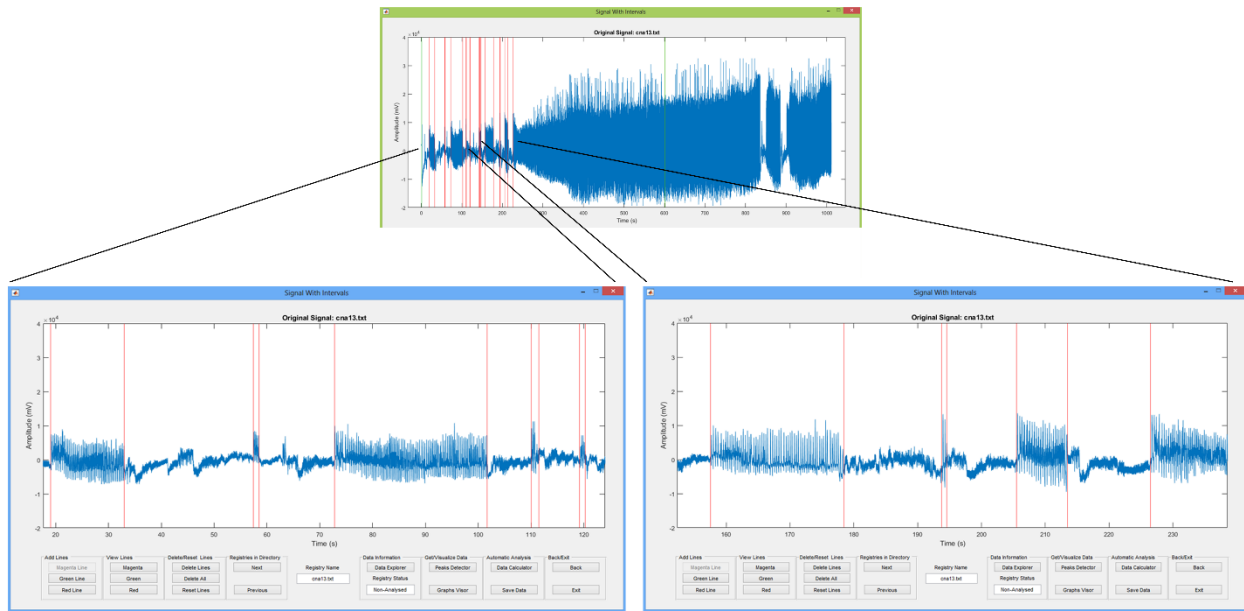


Figura 2.6. Registro electromiográfico (EMG). Las líneas rojas indican los intervalos de bombeo (FVIs) identificados automáticamente por el algoritmo CNNt2. Las dos ventanas inferiores muestran un acercamiento a algunos de estos intervalos.

2.A.4 Discusión

El desempeño más bajo o de menor exactitud (**Accuracy**) de los modelos de entrenamiento con el conjunto de datos *Raw* evidencia que las características esenciales de la señal son menos reconocibles sin su transformación al tipo de datos *FFT*. Con frecuencia, los modelos más complejos y eficaces como las CNNs, son capaces de aprovechar los datos combinados *Raw-FFT* para obtener un reconocimiento más efectivo de las características y lograr mayor exactitud en el proceso de clasificación. Sin embargo, una arquitectura más compleja del modelo no necesariamente implica mejor desempeño, como lo demuestran las combinaciones de CNNs con modelos de RFs, GBs y SVMs, que además requieren más recursos computacionales para su entrenamiento. Las CNNs tienen un margen enorme de configuración debido a la gran cantidad de hiperparámetros que poseen, de modo que con una estructura y configuración adecuadas, una arquitectura profunda de este tipo de modelos puede ofrecer resultados muy interesantes, e incluso un tiempo de entrenamiento relativamente bajo. Ese es el caso del modelo CNNt2, que a pesar de tener un número mayor de capas y utilizar dos tipos de datos para

las entradas, logra tener un desempeño similar al modelo CNNt1, pero con un costo computacional considerablemente más bajo. Adicionalmente, la CNNt2 permite utilizar transferencia de aprendizaje, que es una técnica típica de modelos de aprendizaje profundo en la que se utiliza un modelo pre-entrenado en una tarea como punto de partida para entrenarse en otra tarea. De esta forma, no es necesario entrenar el modelo desde cero para aprender patrones asociados a EMGs realizados con nuevos tratamientos, o con especies diferentes de insectos en nuevas series experimentales. Esta característica es relevante en parte porque no se necesita disponer de una base de datos muy grande para entrenar el modelo en la nueva tarea, a diferencia del caso en el que se tenga que entrenar un nuevo modelo para resolver específicamente dicha tarea. No obstante, MyoRec permite al usuario entrenar un modelo adicional siempre que sea requerido, y la opción para la transferencia de aprendizaje estará disponible en un futuro próximo.

El uso del IEVI en diferentes tipos de EMGs, reveló que la CNNt2 identifica y delimita con bajo porcentaje de error las FVIs en EMGs apetitivos y con bajos niveles de ruido (análisis cualitativo). Rara vez clasifica de forma errónea las ventanas temporales cuando el insecto se alimenta de forma estable y regular durante un tiempo prolongado. Esto significa que normalmente no serán necesarias correcciones por parte del investigador en esos casos. Sin embargo, la efectividad del IEVI desciende en condiciones opuestas, es decir, con EMGs aversivos y con una baja relación señal/ruido. Las soluciones aversivas hacen que el animal se muestre inquieto durante el proceso de alimentación, promoviendo la ausencia de patrones o perfiles de señal regulares y fácilmente identificables. El insecto no solo succiona el alimento de forma errática e interrumpida, con intensidades variables y en cortos períodos de tiempo, sino que además produce numerosas variaciones de voltaje claramente visibles en los EMGs, que no están relacionadas a la actividad de las bombas o músculos succionadores. Estas variaciones son producidas por el movimiento y el roce con la membrana de látex de diferentes partes de su cuerpo, como patas, antenas, probóscide, entre otras, que desfavorecen la relación señal/ruido. Por ejemplo, pueden manifestarse picos de voltaje similares en intensidad y forma a los producidos por la actividad de las bombas, tanto en las FMs como en los períodos de inactividad alimenticia. Esta situación puede confundir al clasificador y provocar falsos positivos. Otros factores relacionados

a la calibración del montaje experimental, como interferencias de ruido electromagnético, también generan alteraciones de la señal que dificultan la correcta identificación y delimitación de las FVIs.

A pesar de la exactitud alcanzada por la CNNt2 (97,4 %), el número de falsos positivos después del proceso de clasificación suele estar por encima del nivel tolerable. Esto se debe a que solo basta un número muy reducido de errores de clasificación para que el número de FVIs detectado por el IEVI sea mucho mayor al verdadero. Sin embargo, el algoritmo de refinación aplicado tras la obtención del vector de predicciones por parte de la CNNt2 contribuye de forma notable a reducir el número de errores. El algoritmo corrige el problema casi en su totalidad para EMGs apetitivos, pero es probable que el investigador tenga que realizar un número relativamente mayor de alteraciones en los resultados con EMG aversivos. Dependiendo de la cantidad de FVIs presentes en un EMG apetitivo, su identificación y delimitación en forma manual puede tomarle a un investigador alrededor de diez segundos a cinco minutos. En un EMG aversivo el período puede extenderse con frecuencia a 30 minutos o más. No obstante, en general la definición automática de las FVIs tanto en registros apetitivos como aversivos mediante el IEVI, reduce el tiempo que se necesita para efectuar los análisis.

PARTE 2: Desarrollo de los Métodos de Cálculo de MyoRec

2.B.1 Introducción

En la medición y cuantificación de variables relevantes como descriptoras del comportamiento de alimentación, tales como por ejemplo: el número y la frecuencia de succiones o bombeos que realiza un insecto, entre otras, el investigador debe realizar varias anotaciones y operaciones aritméticas manuales o ajustar parámetros para cada registro de EMG. Esto hace la tarea notablemente laboriosa y con un gran consumo de tiempo y energía, especialmente cuando los cálculos se hacen en múltiples EMGs. Esta situación ocurre incluso con el uso de la mayoría de los programas actuales para el análisis de este tipo de datos. Las herramientas computacionales muchas veces son específicas para cada tipo de insecto. Por lo tanto, el desarrollo de un módulo de MyoRec que incremente el grado de automatización del cálculo de estas y otras variables, tras el proceso de identificación y delimitación de comportamientos (Parte A), contribuye a agilizar el análisis de los EGMs. En este capítulo se describe el desarrollo de dos aplicaciones modulares para MyoRec. La primera, denominada **Detector de Picos** (DP), realiza el cálculo de parámetros de forma semiautomática en cada FVI (Fase de Verdadera Ingestión), permitiendo al investigador realizar correcciones al número de picos (*i.e.*, bombeos) y sus amplitudes y observar sus efectos en los valores de las variables. Aunque esta aplicación permite al investigador supervisar los cálculos por cada FVI, implica un proceso de análisis más lento que el alcanzado por el **Calculador de Datos** (CD). El CD utiliza una serie de parámetros fijos para realizar estos cálculos de forma totalmente automática, abriendo la posibilidad de analizar múltiples EMGs en un período de tiempo significativamente corto, al compararse con el DP o con cualquier procedimiento manual.

Para manejar los tres módulos principales que conforman MyoRec (IEVI, DP y CD), se desarrolló una interfaz gráfica que incorpora adicionalmente otras aplicaciones con paneles y opciones especializadas. Esto permite al investigador realizar con MyoRec la mayoría de las

tareas asociadas al análisis de EMGs, es decir, organización, almacenamiento, visualización, análisis y comparación de la información, sin que necesite recurrir a programas externos.

2.B.2 Materiales y métodos

2.B.2.1 Detector de Picos

Una vez que se establecen las FVIs del EMG, el cálculo de variables se puede realizar mediante el **Detector de Picos** (DP). Su tarea consiste en detectar y contar los picos de voltaje presentes en las FVIs utilizando el método *Downward Zero-Crossings Smoothed First Derivative* (O'Haver, 2023). Este método considera una muestra (valor puntual de la señal) como un pico de voltaje verdadero si reúne las siguientes tres condiciones: 1) excede un umbral de voltaje preestablecido, denominado Factor de Amplitud (af) (**Figura 2.7**), 2) la primera derivada suavizada de la señal, evaluada en dicha muestra, es cero y cambia su signo en las muestras aledañas (**Figura 2.8**), 3) la segunda derivada supera un valor umbral preestablecido, denominado umbral de pendiente (st) (**Figura 2.9**).

El DP tiene dos modos de funcionamiento para la búsqueda de picos que varían fundamentalmente en la forma cómo se elige el af. El modo “Umbral de Señal Completa (USC)”, descarta los picos de voltaje que tienen una amplitud menor al valor af a lo largo de la FVI. En cambio, en el modo “Umbral de Señal Dividida”, el DP segmenta la FVI en ventanas temporales de 500 muestras de duración (2 segundos) y determina automáticamente un valor umbral semejante al af para cada una. Esto se consigue tras asignar el 20 % del valor medio de la amplitud de los picos dentro de la ventana temporal al parámetro af. El valor del umbral es $af \cdot A$, donde A es el valor promedio de las alturas de los picos a partir del cuantil 70 de su distribución, es decir, el 30 % de los picos más altos. Al definir el umbral de esta forma se evita la necesidad de corregir manualmente los picos identificados en aquellas FVIs de larga duración, y con amplitudes variables en los picos. La eficiencia del proceso se incrementa y disminuye el error de clasificación.

El proceso de suavizado, al aplicar el método *Downward Zero-Crossings Smoothed First Derivative* reduce la amplitud de los picos, por lo que se definió el hiperparámetro **Factor de**

Varianza (wf) que restablece la altura original de los picos de voltaje. Para ello establece un intervalo alrededor de la muestra identificada como pico y la sustituye por aquella con el valor de voltaje más alto si la hubiese. Para reducir la sensibilidad al ruido o de picos artefactuales también se definió el **Factor Horizontal** (hf). Este hiperparámetro también establece un intervalo alrededor de un pico pero, en lugar de considerar todas las muestras presentes en el intervalo, considera sólo aquellas que también hayan sido identificadas como picos. De este modo cambia la muestra por la del pico más alto en el intervalo. Si el pico alrededor del cual se forma el intervalo es el más alto, no se modifica el estado de la muestra, *i.e.* si hay más de un pico en el intervalo, solo prevalece el de mayor amplitud.

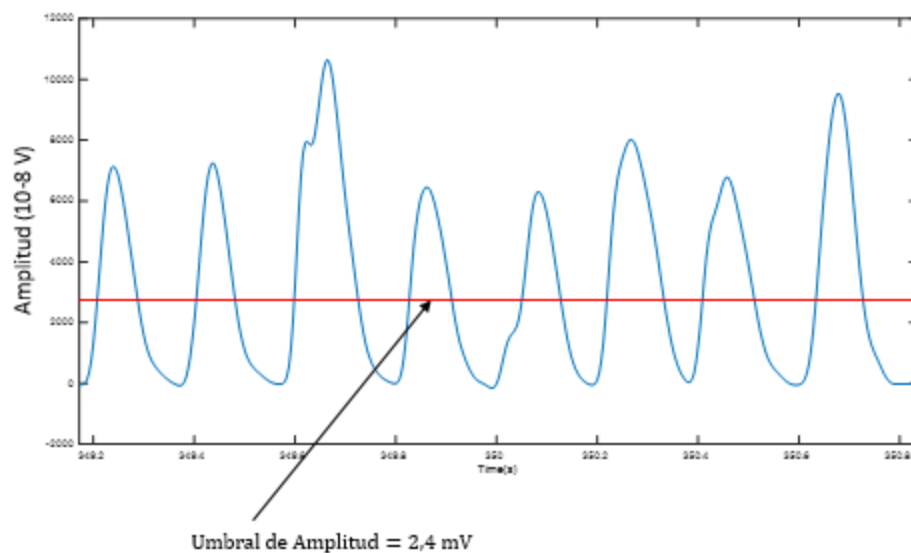


Figura 2.7. Representación ilustrativa de una señal y un valor umbral de voltaje (línea roja). Si el punto en cuestión tiene un valor de voltaje mayor al umbral habrá cumplido con el primer requisito necesario para calificar como pico o bombeo.

Curva Original (Azul) y su Primera Derivada (Roja)

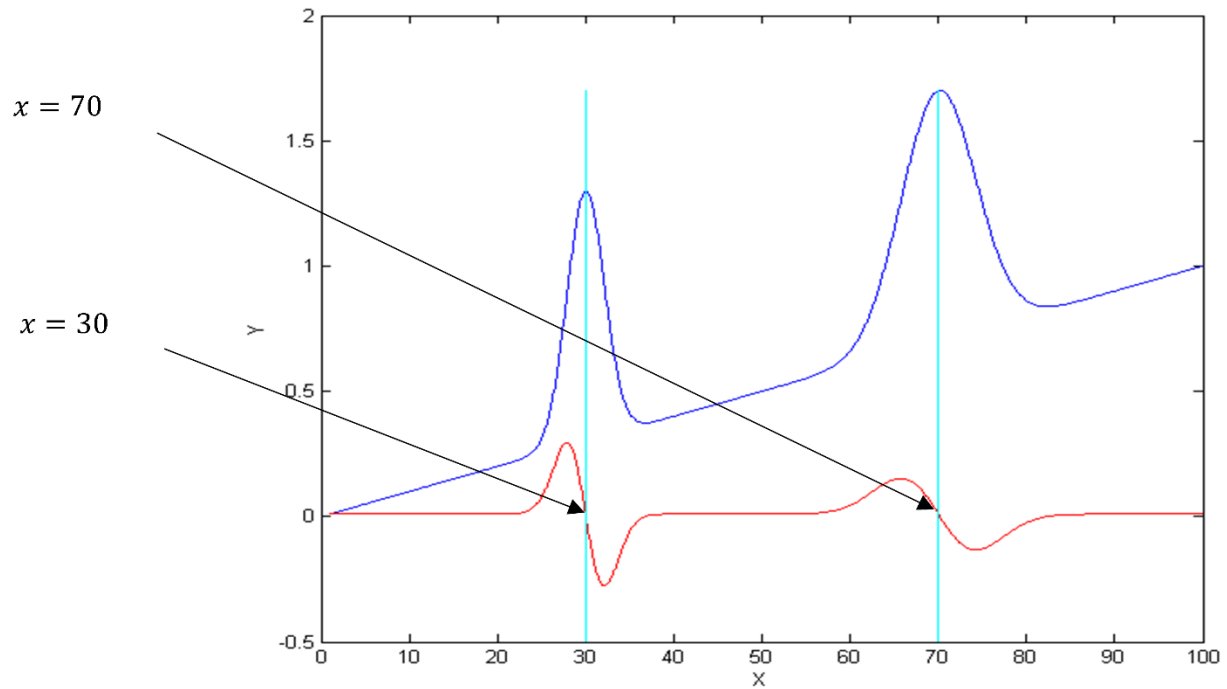


Figura 2.8. Representación ilustrativa de una señal (azul) con su derivada (roja) y los puntos en los que esta cambia de signo, lo cual satisface el segundo de los requisitos para que sean considerados picos de voltaje o bombeos.

Curva Original (Azul) y su Primera Derivada (Roja)

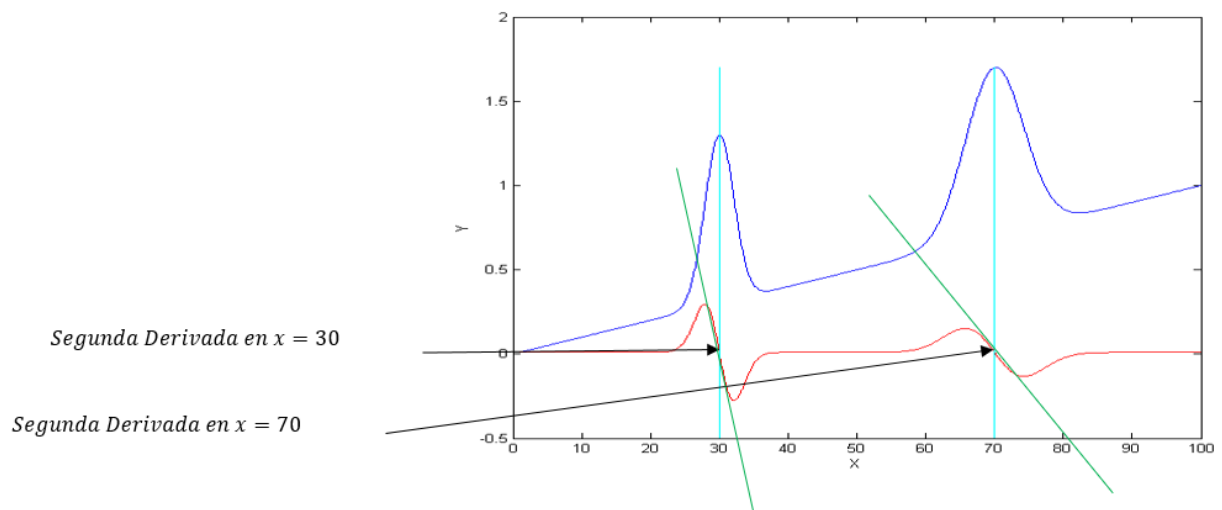


Figura 2.9. Representación ilustrativa de una señal (azul) con su derivada (roja) y los puntos en los que esta cambia de signo (cian), lo cual satisface uno de los requisitos para que sean considerados picos de voltaje o bombeos. El tercer requisito se satisface si la segunda derivada presenta una pendiente superior a un valor umbral.

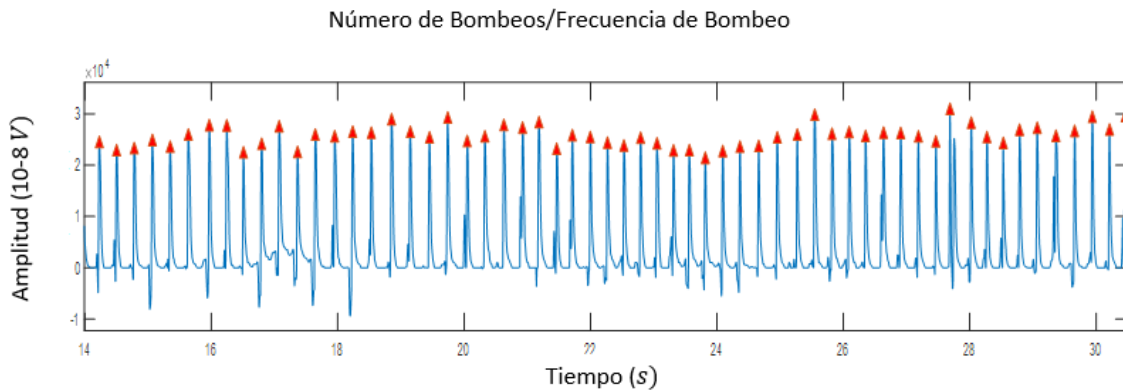


Figura 2.10. Detección y cuantificación de los picos de voltaje mediante el método *Downward Zero-Crossings Smoothed First Derivative* (O'Haver, 2023).

2.B.2.2 Calculador de Datos

El **Calculador de Datos** (CD) es una aplicación que tiene el objetivo de realizar el cálculo de las variables de la forma más rápida y autónoma posible. El algoritmo elimina la media de la señal, la segmenta en ventanas temporales solapadas parcialmente y aplica en cada una la transformada de Fourier (**Figura 2.11**), lo que se conoce como el método Welch. Cada una de las ventanas contiene 1024 muestras, longitud de ventana (wl), por defecto, pero puede ser modificada por el usuario, es decir, es un hiperparámetro. Otro hiperparámetro es el solapamiento (ovl), el cual modifica el grado de superposición de las ventanas. Su valor se estableció por defecto en un 90 %. Esto permite ofrecer un balance entre la definición de los cambios frecuenciales en el tiempo y la resolución frecuencial. Por convención se asume que la frecuencia fundamental de la señal en cada ventana, coincide con el valor de la frecuencia de bombeo y está representada por el pico más alto en el dominio de frecuencias. Debido a que la frecuencia de bombeo en *R. prolixus* presenta en la enorme mayoría de los casos valores mayores

a los 2 Hz y menores a los 10 Hz, el algoritmo evalúa la frecuencia del pico más alto en el intervalo [2 - 10] Hz en cada ventana. No obstante, estos límites pueden cambiar si se estudian otras especies de triatominos u otros insectos. Las FVIs están compuestas generalmente por numerosas ventanas temporales, por lo que los valores de frecuencia registrados en cada una se promedian. Para las FVIs con menos de 1500 muestras (6 segundos de duración), la frecuencia de bombeo se calcula promediando el número de picos detectados por el DP y el método Welch. Esto mejora la estimación de la frecuencia cuando existe un número bajo de picos en las FVIs. Por otra parte, las amplitudes de tales picos se registran automáticamente mediante la aplicación del DP con hiperparámetros fijos para todas las FVIs del EMG. A partir de la frecuencia de bombeo y las amplitudes, el algoritmo calcula el resto de las variables mediante operaciones matemáticas simples (se describen en el Capítulo 3).

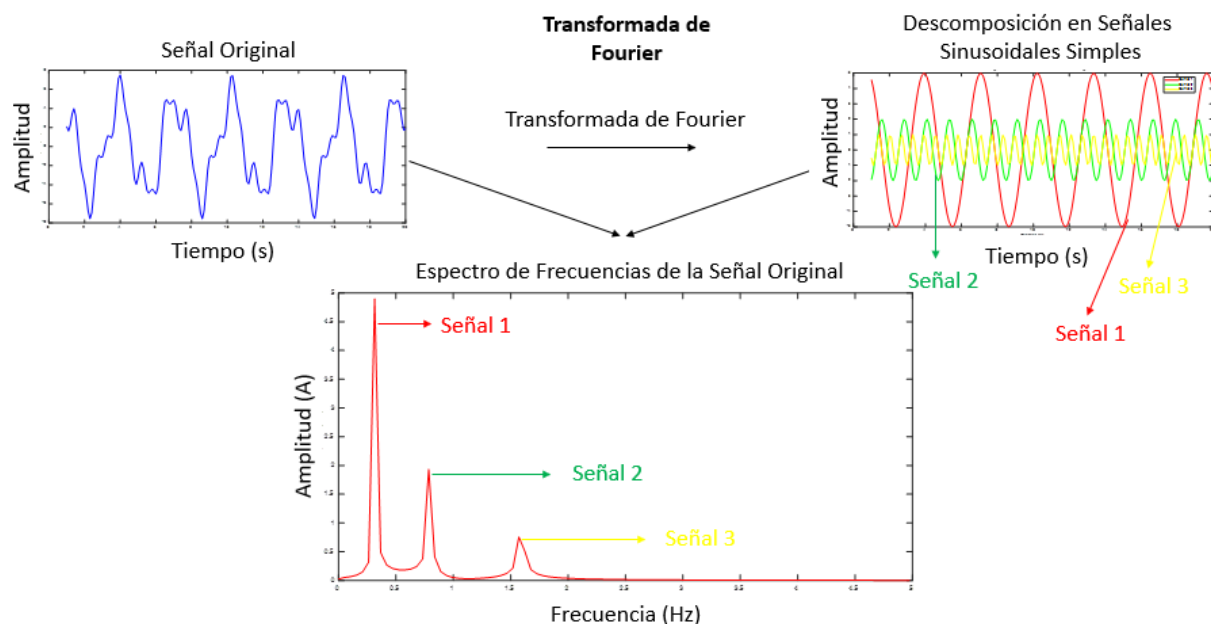


Figura 2.11. Descomposición de una señal en ondas sinusoidales simples de diferente frecuencia al aplicar la transformada de Fourier. **A** – Señal compleja formada por tres señales sinusoidales simples con diferentes frecuencias e intensidades. **B** – Señales sinusoidales simples que al sumarse dan como resultado la señal en **A**. **C** – Picos de intensidad en el dominio de frecuencias pertenecientes a las señales sinusoidales simples en **B**, obtenidas al aplicar la transformada de Fourier a la señal compuesta de **A**.

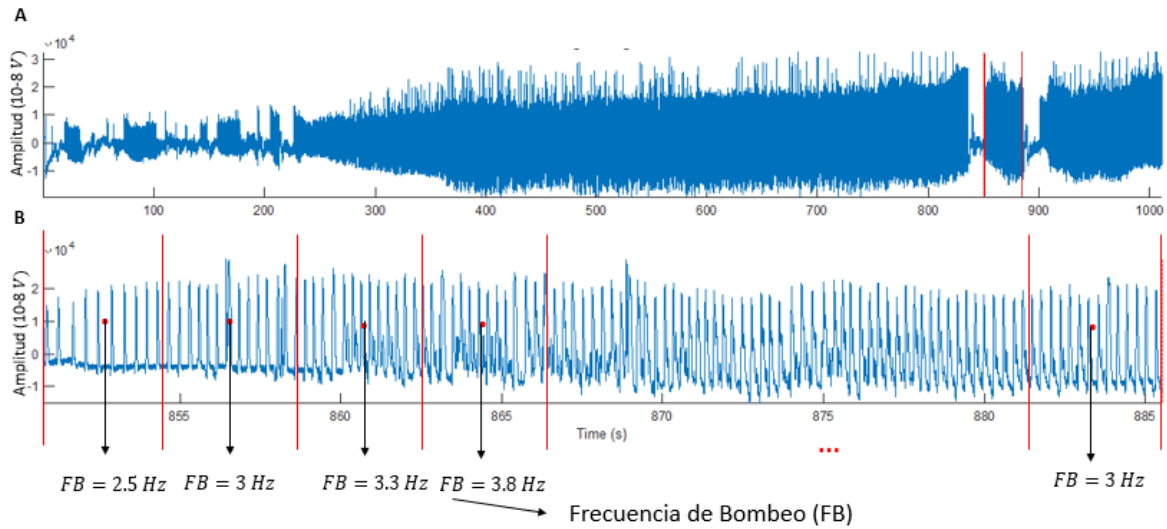


Figura 2.12. Aplicación de la Transformada de Fourier a las ventanas temporales que dividen la señal. **A** – Señal original. Las líneas rojas representan los límites de una FVI. **B** – Las líneas rojas representan las ventanas donde se aplica la Transformada de Fourier.

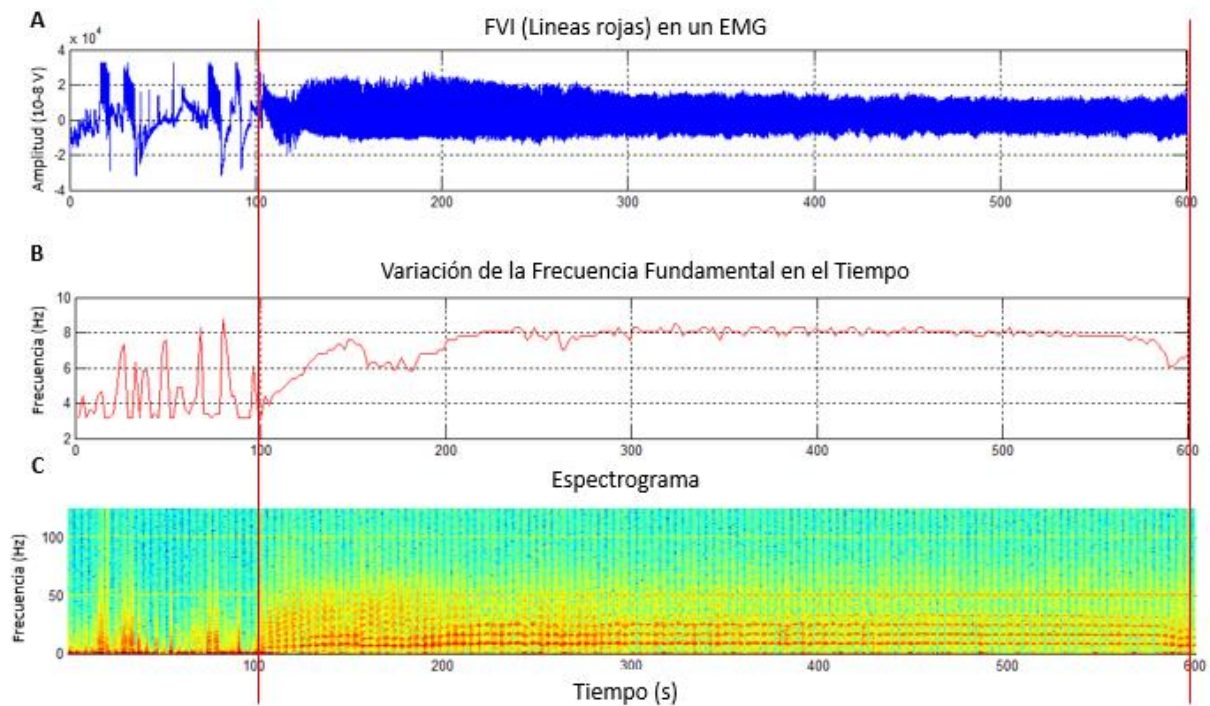


Figura 2.13. Aplicación de la Transformada de Fourier en ventanas temporales solapadas para obtener la frecuencia fundamental (Método Welch). **A** – FVI delimitada con líneas rojas en la señal original. **B** – Variación de la Frecuencia Fundamental en el Tiempo. **C** – Espectrograma de la señal.

2.B.2.3 Efectividad y Eficiencia de Cálculo

Se analizaron un total de 100 EMGs utilizando las aplicaciones **DP** y **CD** para comparar sus resultados. Para determinar la efectividad del algoritmo del **CD** al analizar EMGs, se calculó el error porcentual entre los resultados producidos tanto por el **DP** como por el **CD** en las siguientes variables: FB (frecuencia de bombeos, calculada a partir del número de bombeos durante la FVI), FBC (frecuencia de bombeos durante la totalidad del período de contacto o período en que el insecto mantuvo sus estiletes en el interior del alimentador) y A (amplitud media de los picos o bombeos). Para ello, se utilizó la siguiente ecuación:

$$PE_i = [|PD_i - DC_i|/PD_i]100, \quad i = 1, \dots, N$$

donde PE_i es el i-ésimo error porcentual, PD_i es el i-ésimo elemento de la variable calculada con el **DP**, mientras que DC_i es el i-ésimo elemento de la variable calculada con el **CD**. Los EMGs se dividieron en tres categorías para evaluar la precisión del algoritmo según el número de bombeos (NB) presentes en cada uno. El grupo corto: de 10 a 100 ($N = 41$), grupo medio: de 100 a 500 ($N = 20$) y grupo grande: más de 500 ($N = 39$). Los EMGs con menos de diez picos en total fueron excluidos del análisis. El CD se configuró con los siguientes hiperparámetros: $wl = 1024$ muestras y $ovl = 960$ muestras, $af = 0,05$, $st = 4,6094e-08$, $sw = 3$, $pg = 2$, $sty = 4$, $wf = 2$, $lwf = 4$ y $hf = 9$ (**ver Apéndice B**).

La eficiencia de los algoritmos se determina de acuerdo al tiempo que tarda el investigador en calcular las variables de un registro completo. Esto resulta muy variable en el caso del **DP** debido a que implica la supervisión manual de todas las FVIs de la señal para agregar o eliminar los picos identificados por el algoritmo. Adicionalmente es posible que se requiera una calibración de los hiperparámetros en diferentes FVIs. En el caso del **CD** el cálculo de variables es completamente automático a lo largo del EMG, por lo que el tiempo que se requiere depende solo de la rapidez del algoritmo.

2.B.3 Resultados

2.B.3.1 Interfaz Gráfica

El programa MyoRec fue creado principalmente en el lenguaje de programación Octave, y posee algunas partes desarrolladas en Python y C++. Está compuesto por múltiples aplicaciones que brindan al usuario toda clase de opciones y funcionalidades que se necesitan para procesar los datos. Las aplicaciones se dividen en tres categorías: aplicaciones principales, destinadas a brindar todas las funciones necesarias para almacenar, analizar, visualizar y comparar datos de EMGs (cuadrados verdes en la **Figura 2.14**), aplicaciones complementarias, que brindan información y procesos adicionales para facilitar las tareas de las aplicaciones principales (cuadrados azules en la **Figura 2.14**), y aplicaciones de diálogo y configuración, que en general permiten al usuario establecer hiperparámetros y tomar decisiones sobre el modo en el que se van a procesar, almacenar o visualizar los datos (cuadrados rosados en **Figura 2.14**).

En las secciones subsiguientes se describen brevemente las aplicaciones de la interfaz gráfica. Sin embargo, para mayor claridad se puede encontrar información adicional explicada en formato de video siguiendo este link: https://github.com/hecsalms/MyoRec_GUI.

2.B.3.1.1 Aplicaciones Principales

Las aplicaciones principales incluyen ocho ventanas llamadas Panel Principal (MP), Panel de Control de Datos (DCP), Visualizador de Señales (SV), Explorador de datos (DE) (**Figura 2.15A**), Interfaz de Control de Procesos (CPI), Detector de Picos (PD), Calculador de Datos (DC) y Visor de Gráficos (GV) (**Figura 2.15B**). El MP permite cargar los archivos de datos crudos en formato “.txt”, y visualizar la señal que contienen (EMG) en la aplicación SV. Los archivos de datos que han sido generados por el programa tras el proceso de análisis tienen formato “.mat”. A diferencia de los archivos “.txt”, los datos contenidos en estos archivos se visualizan en la aplicación CPI, que muestra la señal con los límites marcados para las FVIs (líneas rojas), los PCs (líneas verdes), y el tiempo de registro (líneas rosadas). Desde el MP se pueden realizar gráficos de diagramas de cajas y de evolución temporal para los datos de las variables contenidos en los archivos “.mat”, y también es posible acceder a la aplicación DCP, que permite inspeccionar las tablas de datos

asociadas a estos archivos. Específicamente, la aplicación DCP muestra la información de los archivos “.mat” en dos tablas. La tabla superior contiene valores acumulativos y promedios de las variables, mientras que la tabla inferior muestra los valores que las variables adoptan para cada FVI o PC según el caso. Esta información varía según la variable seleccionada en una lista con los nombres de cada una. Los nombres de los registros y las posiciones de las filas se pueden editar en la tabla superior. Las filas de datos se pueden eliminar, reemplazar o agregar al cargar otros archivos. El archivo “.mat” se puede actualizar con los datos modificados en las tablas y también se pueden crear archivos “.xlsx” y “.txt” para guardar la información, aunque en el último caso, sólo los datos de la tabla superior se almacenan. Desde la DCP se puede acceder a una aplicación complementaria llamada “Información de hiperparámetros (II)”, que muestra las coordenadas de los picos y los parámetros de configuración del DP o el CD (según el caso) que se usaron para identificarlos en cada FVI.

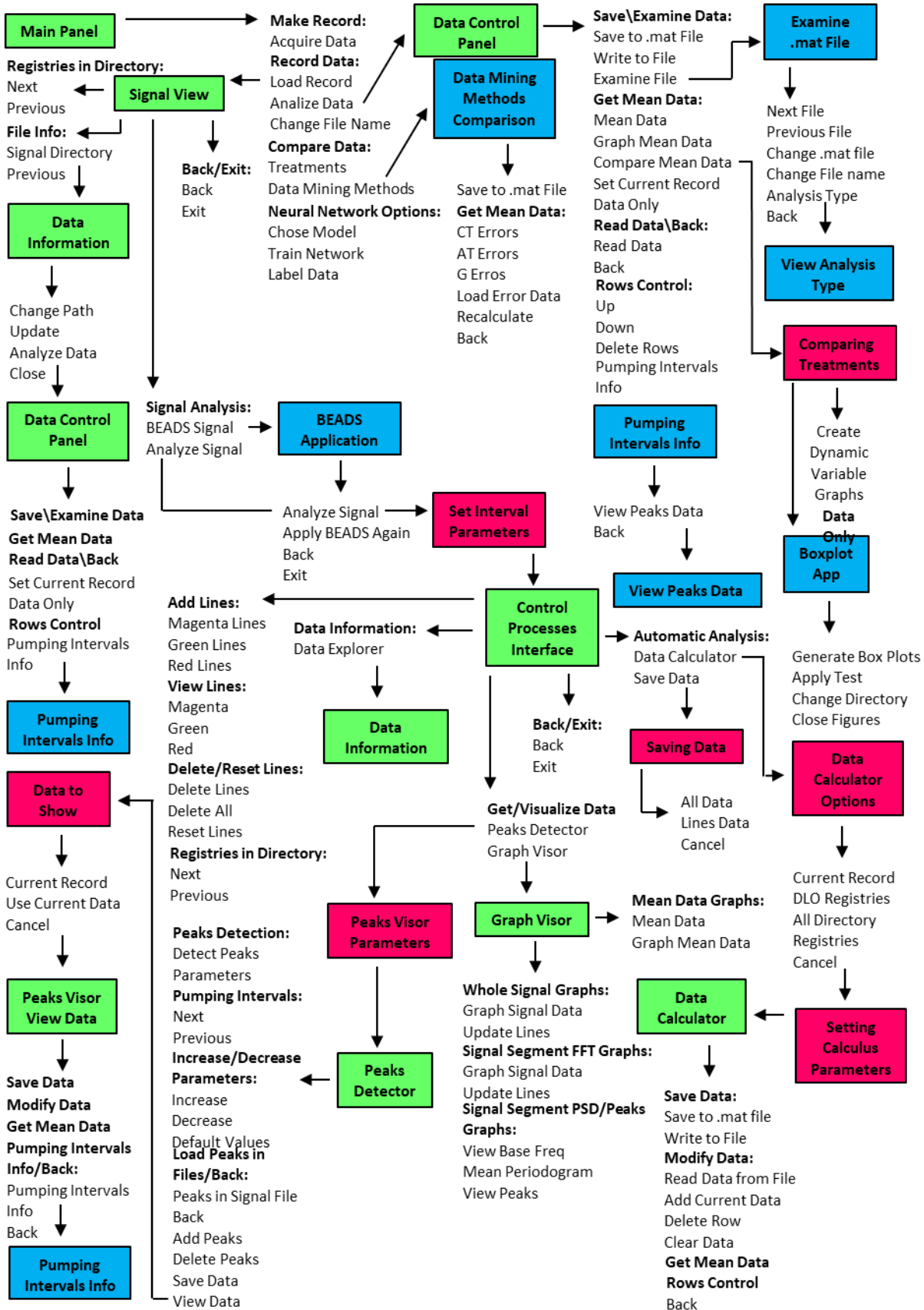


Figura 2.14. Aplicaciones y funciones que componen MyoRec. Los cuadrados de colores indican los nombres de las aplicaciones. Los cuadrados verdes indican las aplicaciones principales, mientras que los azules y rojos se refieren a aplicaciones complementarias y de enlace respectivamente. Los textos con puntas de flecha muestran las funciones activadas por botón en la aplicación de donde sale la flecha. Los nombres de los paneles que contienen los botones en la aplicación de la ventana se indican en negrita. Las funciones que activan una aplicación en particular se indican con una flecha que sale de su nombre hacia el cuadrado correspondiente. Aunque algunas aplicaciones pueden tener botones con la misma función, a veces se omiten para evitar la redundancia.

Al cargar el archivo crudo (en formato .txt) de un EMG, se abre la aplicación SV que permite visualizar la señal. El usuario puede acercarse, alejarse y desplazarse a lo largo de los ejes para observar más detalles de la misma. Desde la interfaz de la SV se pueden cargar archivos de datos que también se encuentren en el directorio actual mediante la aplicación DE. También se pueden calibrar los hiperparámetros del algoritmo BEADS al observar sus efectos sobre la señal con la aplicación complementaria (BA). Al presionar el botón de analizar datos en la SV o la BA, se abre una ventana de configuración (*Set Hyperparameters*, SHP) de hiperparámetros para el IEVI, aunque su calibración se recomienda solo para usuarios avanzados. Al aceptar esta configuración se visualiza posteriormente el EMG en la CPI, donde aparecen opciones adicionales relacionadas a la identificación y delimitación de los eventos de alimentación, a la visualización de los datos y al cálculo de variables. Esta aplicación puede considerarse como la más importante, porque proporciona acceso a casi todas las aplicaciones relacionadas con el análisis, visualización y almacenamiento de los datos. Una de estas aplicaciones es el DE, que posee un explorador de directorios para cargar nuevos archivos de datos y muestra en una tabla las coordenadas de los límites de los diferentes tipos de eventos marcados sobre la señal. Al mismo tiempo muestra la cantidad de períodos de contacto (PCs), FVIs y el número de eventos de alimentación, lo que ayuda a prevenir errores por parte del usuario.

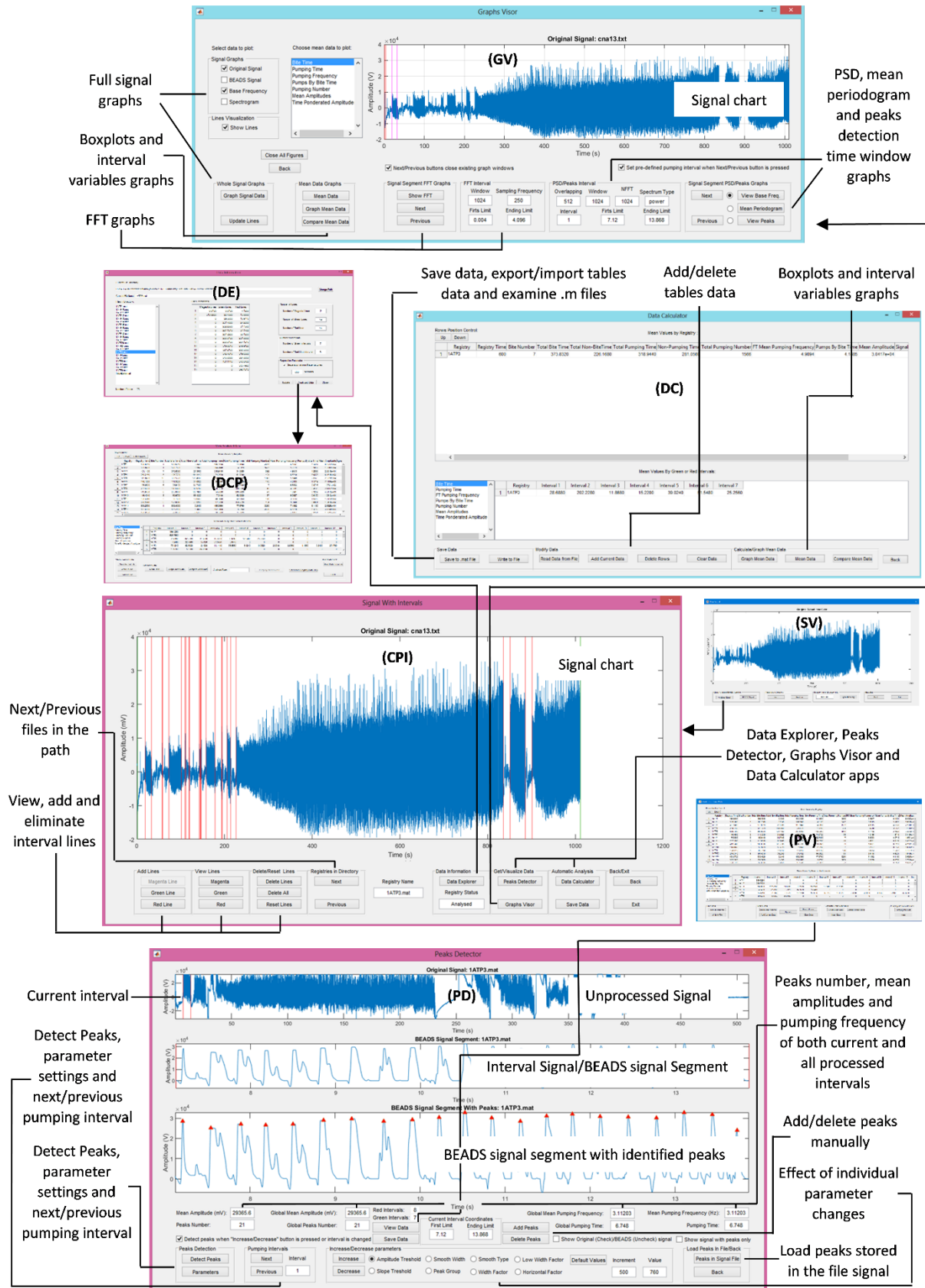


Figura 2.15. Aplicaciones principales y complementarias que el usuario puede activar desde la aplicación principal MP. El usuario accede a la CPI desde la SV o directamente desde la MP cargando un archivo .mat. Las flechas que salen de un botón indican las aplicaciones que activa, en algunos casos, a través de una aplicación de diálogo (no se muestra). **A** – Aplicaciones MP, SV, CPI, DE, DCP, DE, BS y DMM, sus conexiones y principales funciones. **B** – Aplicaciones CPI, DP CD, DCP, DE, GV, SV y PV, sus conexiones y principales funciones.

Una vez que los eventos de alimentación han sido marcados en la CPI, se deben calcular las variables que cuantifican la información sobre el comportamiento alimenticio del animal. Para ello, la CPI ofrece acceso a las aplicaciones DP o CD, que realizan los cálculos con dos niveles de automatización diferentes, como se ha descrito en forma previa (ver secciones **2B.2.1** y **2B.2.2**). Si el investigador decide utilizar el DP para el análisis, se le ofrece configurar los hiperparámetros del algoritmo antes de comenzar, aunque con frecuencia la configuración por defecto resulta adecuada. La interfaz del DP muestra tres canvas de forma predeterminada que contienen, en el primer canvas, la señal total sin procesar con la FVI actual demarcada con líneas verticales rojas. En el segundo, la FVI con el preprocesamiento BEADS, y en el tercer canvas, el mismo segmento de señal pero con los picos identificados. El *zoom* y las opciones de desplazamiento pueden activarse en cualquiera de los canvas, pero esto resulta especialmente útil en el tercero, cuando se está analizando una FVI de larga duración. En la parte inferior del conjunto de canvas, existen numerosas opciones que facilitan las operaciones relacionadas a la identificación de los picos y el cálculo de las variables. Aunque se muestra información parcial de algunas de ellas, como el número de picos, sus amplitudes y la frecuencia de bombeo, la interfaz del DP también permite el acceso a una aplicación muy similar a la DCP, llamada *View Data* (VD). En ella se puede observar toda la información relacionada a las variables a medida que se analizan las FVIs. La aplicación DP también hace posible observar los picos identificados en un análisis previo al cargar los datos registrados en un archivo “.mat”. Esto permite evaluar la calidad del análisis y hacer modificaciones si es necesario, sin tener que realizar el análisis desde cero. El algoritmo que identifica los picos en las FVIs puede ajustarse mediante la configuración de los hiperparámetros o de forma manual utilizando las funciones de agregación y eliminación de picos.

Si el investigador escoge la aplicación CD para el análisis desde el CPI, debe configurar, de manera similar al DP, los hiperparámetros correspondientes, indicando si el análisis se realiza sobre un solo EMG o sobre un conjunto de ellos. La interfaz del CD es similar a la del DCP o el VD, y mostrará los datos en las tablas a medida que el, o los EMGs sean analizados.

Otra aplicación importante a la que se accede desde la CPI es el GV. Su función principal es proporcionar información gráfica sobre los datos del EMG. Existen dos categorías de visualización de datos: los gráficos completos, que cubren toda la longitud de la señal, y los gráficos parciales, que visualizan sólo un segmento de la señal, e.g., una FVI. Los gráficos completos generalmente se muestran en ventanas emergentes que pueden incluir varios canvas con gráficos del EMG original, el EMG procesado con BEADS, la frecuencia basal en el tiempo y un espectrograma. También entran en esta categoría los gráficos de evolución temporal de las variables (los cuales también pueden visualizarse desde otras aplicaciones como MP, DCP, VD y DC). Los gráficos parciales se usan principalmente para facilitar la configuración de hiperparámetros en las aplicaciones DP y CD, al mostrar visualmente los efectos de cada configuración en la señal. Como ejemplo, el investigador puede definir los límites del segmento del EMG en el que aplica la Transformada Rápida de Fourier (FFT) y observar el resultado en una ventana emergente. Esta ventana muestra tres canvas que contienen la señal original y los límites donde se aplica la FFT (indicados por líneas verticales rojas), la señal preprocesada con BEADS, y el espectro de frecuencias respectivamente. Otras opciones, como “Ver Frecuencia Basal” (algoritmo en el CD) y “Ver Picos” (algoritmo en el DP), muestran adicionalmente el valor de la frecuencia basal, la longitud del intervalo, la amplitud media, el número de picos, entre otros valores.

2.B.3.1.2 Aplicaciones Complementarias

Las aplicaciones complementarias están formadas por el *Comparador de Métodos de Extracción de Datos* (DMMC), *Aplicar BEADS* (BA), *Información de FVIs* (PII), *Ver Datos de los Picos* (VPD), *Examinar Archivos .mat* (EmF), *Ver Tipo de Análisis* (VAT) y la aplicación *Diagrama de Cajas* (BPA). Algunas de ellas, como BA, PII, VPD y VAT ya se han descrito previamente, mientras que DMMC, EmF y BPA merecen una breve descripción. El Comparador de Métodos de Extracción de

Datos (DMMC) se abre desde la aplicación MP y se utiliza para comparar los resultados del algoritmo del CD con los obtenidos por el DP, tras analizar los EMGs de forma automática y supervisada, respectivamente. Debido a que el análisis de los datos con el DP es supervisado por el investigador de forma detallada, se utiliza como referencia para calibrar los hiperparámetros del CD para que su algoritmo obtenga los resultados más cercanos a los del DP. Una tabla superior muestra los errores porcentuales de tres variables (FB, FBC y A). Otra tabla en la parte derecha de la ventana muestra el tratamiento al que pertenecen los datos, mientras que en la parte inferior se muestra el error porcentual medio de cada variable. La aplicación *Examine .mat File* (EmF) sirve para inspeccionar los datos almacenados en un archivo “.mat”, y por último, la aplicación *View Analysis Type* (VAT) muestra el tiempo de registro del EMG y el método que fue utilizado para analizarlo (*Peaks Detection* o *Welch Method*) según la información en el archivo “.mat” correspondiente. Finalmente, la aplicación *Boxplot* (BA) permite hacer diagramas de cajas sobre las variables contenidas en este tipo de archivos que se encuentran en el mismo directorio. Para cada gráfico se realiza por defecto una prueba estadística de Kruskal Wallis.

2.B.3.1.3 Aplicaciones de Configuración

Las aplicaciones de diálogo y configuración se utilizan para mostrar instrucciones de uso, establecer hiperparámetros, configuraciones, enlaces a otras aplicaciones, y toma de decisiones. Están conformadas por *Establecer Hiperparámetros* (SIP), *Comparar Tratamientos* (CT), *Datos para Mostrar* (DTS), *Guardar Datos* (SD), *Opciones del Calculador de Datos* (DCO), *Hiperparámetros para la Identificación de Picos* (PVP) y *Configuración de Hiperparámetros para el Cálculo de Variables* (SCP) (cuadrados rosados en la **Figura 2.14**).

2.B.3.2 Estructura de los Archivos de Datos

Los datos de los EMG analizados con MyoRec se almacenan en dos tipos de archivos con extensión “.mat”; archivos individuales y colectivos. Los archivos individuales poseen información sobre un solo EMG, mientras que los colectivos contienen datos sobre todos los EMGs que son realizados para un tratamiento. Algunos datos solo se encuentran en los archivos individuales, como las coordenadas de los picos de voltaje que permiten reproducir el análisis posteriormente y realizar modificaciones en caso de ser necesario. Por consiguiente, en el

directorio correspondiente a un tratamiento se encuentran normalmente ambos tipos de archivos, es decir, múltiples archivos individuales y un archivo colectivo. El usuario puede colocar un nombre cualquiera a los archivos colectivos pero siempre deben contener el término 'TableData' para que las aplicaciones de MyoRec los reconozcan.

2.B.3.3 Rendimiento del Detector de Picos y del Calculador de Datos

Las aplicaciones para el cálculo de variables en MyoRec (DP y CD) tienen propósitos diferentes. El DP se diseñó para que el investigador pueda obtener el valor de las variables con el menor número de calibraciones posible, mientras permite supervisar el proceso de identificación de los picos en el EMG para cada FVI y hacer modificaciones cuando se requiera. El algoritmo que detecta los picos tarda solo unos segundos en detectar cientos de picos en una FVI y calcular todas las variables con esa información, lo que hace que el proceso de análisis pueda llevarse a cabo con gran rapidez. Sin embargo, el CD se desarrolló para automatizar completamente el cálculo de variables estableciendo hiperparámetros fijos, lo que permite terminar el análisis de un registro con 3500 picos en aproximadamente cinco segundos. Esto posibilita que en solo tres minutos pueda completarse el cálculo de variables de aproximadamente 15 EMGs de esta categoría con una computadora ordinaria (I7 3230Q, 16 Gb RAM). El proceso será incluso más rápido en EMGs con una cantidad menor de picos o una computadora más potente y rápida.

2.B.3.4 Efectividad del Calculador de Datos

El error porcentual en el grupo reducido fue de 15,6 % para la variable FB y del 12,7 % para la variable AM (**Figura 2.16A**). En el grupo medio fue de 11 % y 14,6 % respectivamente (**Figura 2.16B**), mientras que el grupo grande tuvo los valores más bajos de error, con 1,98 % para la FB y 8,17 % para AM (**Figura 2.16C**). La **Figura 2.16D** muestra el error de FB, FBC y AM incluyendo todos los tipos de EMGs e indica un error de 9,42 % en la FB y 11,41 % para la AM. El error porcentual para la FBC fue muy similar al de la FB en todos los casos.

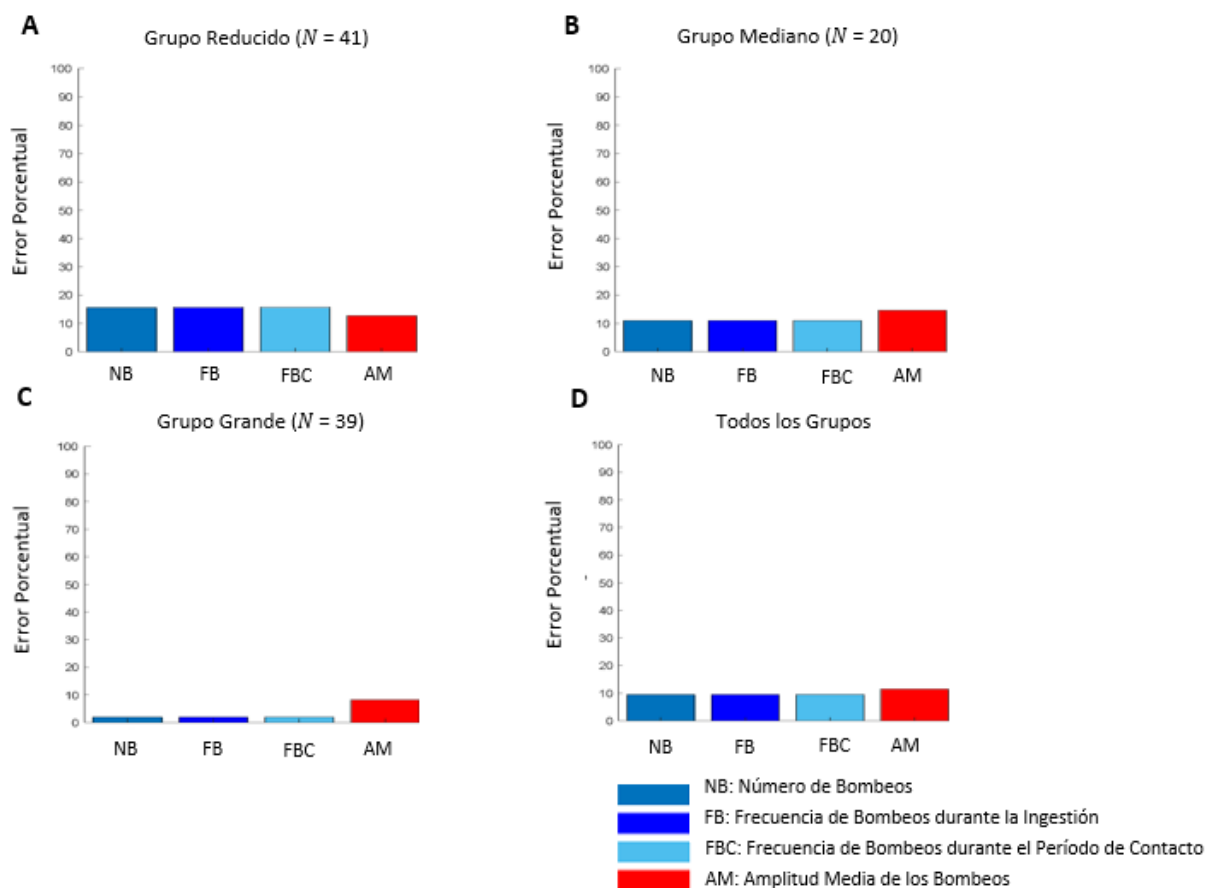


Figura 2.16. Error porcentual entre los valores predichos por el Calculador de Datos y los obtenidos con el Detector de Picos para tres grupos diferentes de EMGs ($N = 100$), basados en la cantidad de bombeos realizados por el insecto durante las fases de verdadera ingestión. **A** - El grupo reducido: 10 – 100 bombeos por registro ($N = 41$), grupo mediano: 100 – 500 bombeos ($N = 20$) y grupo grande: más de 500 bombeos ($N = 39$). Señales con menos de diez bombeos fueron excluidas del análisis. Los valores de los parámetros del método de Welch utilizados fueron $wl = 1024$ muestras y $ovl = 960$ muestras, mientras que los usados para la detección de picos fueron $af = 0,05$, $st = 4.61e-8$, $sw = 3$, $pg = 2$, $sty = 4$, $wf = 2$, $lwf = 4$ y $hf = 9$.

2.B.4 Discusión

La interfaz gráfica de MyoRec simplifica el análisis de los EMGs respecto a los análisis manuales y programas con interfaces gráficas no especializadas. La interfaz de MyoRec ofrece un sistema de ventanas fácil de usar diseñado para proporcionar una amplia gama de opciones que se necesitan para administrar los archivos de datos crudos y los que contienen datos extraídos de los análisis. Proporciona herramientas de edición y corrección en las tres partes principales

del proceso de análisis, como identificación de comportamientos, la detección de picos y el cálculo de variables, permitiendo modificar los resultados de las operaciones automáticas cuando sea necesario, ya sea durante el proceso de análisis o en una sesión futura.

Los servicios automáticos que proporciona MyoRec, como el IEVI o el cálculo automático de variables es una virtud que no poseen muchos de los programas no especializados, y es probable que incluso aquellos que cuentan con servicios similares no funcionen adecuadamente en los EMGs de triatominos. En el caso del IEVI, las imprecisiones de otros algoritmos presentes en otros programas se deben a que los patrones de onda observados en los EMGs con triatominos son significativamente distintos de aquellos con los que fueron entrenados. Esto hace que en algunos casos sean completamente incapaces de detectar las FVIs. Respecto al cálculo de variables, los programas no especializados simplemente no pueden realizar el proceso con el mismo grado de automatización y probablemente requieren de calibraciones de parámetros con mucha frecuencia, no sólo entre diferentes EMGs sino también en un mismo registro. Esto en definitiva ralentiza el proceso de análisis y lo hace más exhaustivo. Con las aplicaciones DP y CD de MyoRec, se consigue eliminar o disminuir significativamente este inconveniente, especialmente con el CD, que permite el cálculo de variables en múltiples EMGs de forma casi instantánea tras presionar un botón. La exactitud del algoritmo del CD para las variables FB y FBC mejora al incrementarse el número de picos (bombeos) en las FVIs de los EMGs. Los errores en grupos de EMGs cortos y medianos que superan el 10 %, podrían ser el resultado de una mayor dificultad para reconocer los picos de voltaje correctos por parte del algoritmo. La transformada de Fourier ofrece resultados menos exactos a medida que se reduce la duración del segmento de la señal sobre el que se aplica, es decir, la FVI. Los EMGs de corta duración suelen poseer numerosas FVIs de corta duración y con un número muy pequeño de picos que con frecuencia no sobrepasa las 10 unidades. Es por eso que en estos casos se promedia el resultado del método Welch con el del algoritmo de detección de picos con hiperparámetros fijos. Sin embargo, estos picos tienen estructuras irregulares y más difíciles de distinguir del ruido y los picos provocados por la bomba faríngea, por lo que el algoritmo no está exento de errores. El error disminuye en los EMGs de categoría media porque aparecen FVIs de mayor duración, con mayor cantidad de picos, con una frecuencia de bombeo más regular, y en consecuencia ésta se determina con más

exactitud. Finalmente, en los EMGs grandes el error es mínimo y los resultados son casi idénticos a los obtenidos con el DP porque las FVIs suelen abarcar casi toda la señal. Esto significa que para registros medios y grandes los beneficios de usar el CD son significativos considerando que puede calcular las variables de decenas de registros en cuestión de minutos con una gran exactitud. En el caso de los EMGs cortos, la frecuencia de bombeo no se puede determinar con mucha exactitud independientemente del algoritmo que se utilice, debido a que la cantidad de datos es reducida. Si intentamos medir la frecuencia de bombeo para una FVI con tres picos de voltaje, simplemente no hay forma de calcularla de forma insesgada. Respecto a la amplitud, no existe una influencia clara del número de picos en la FVI, y la disminución del error porcentual en los EMGs grandes probablemente se debe a que existe una mayor regularidad de los picos y una mayor relación señal/ruido en las FVIs. El error porcentual en las amplitudes puede disminuir escogiendo una configuración de hiperparámetros más adecuada. Sin embargo, en la mayoría de los casos no es necesario obtener las amplitudes con un alto grado de precisión, a diferencia de la FB.

El error porcentual para la FB y la FBC se mantuvo por debajo del 10 % considerando todos los grupos, mientras que la amplitud fue ligeramente mayor a este valor. Esto nuevamente demuestra que es posible determinar automáticamente la información de las variables con un grado aceptable de confiabilidad.

Capítulo 3: Uso de MyoRec en el Análisis del Comportamiento de Alimentación

3.1 Introducción

El sistema gustativo en insectos les permite distinguir entre una fuente de alimento rica en nutrientes, con alto grado de palatabilidad o apetitiva, y una con bajo contenido nutricional o incluso tóxica y por ende aversiva (King y Gunathunga, 2023). Los quimiorreceptores de contacto localizados en el órgano faríngeo de las vinchucas (**Figura I.5**) proveen información sobre la calidad del alimento que ingresa al insecto, por ejemplo la sangre, alimento fundamental de las vinchucas (Pontes et al., 2014). Sin embargo, también ayuda en la detección de compuestos potencialmente tóxicos o nocivos (Barrozo, 2019). Esta detección previa a una ingesta a repleción previene y alerta al insecto de aquello que va a ingerir. En el interior de estos quimiorreceptores de contacto se alojan neuronas sensoriales gustativas (NSGs), las cuales poseen receptores moleculares en sus dendritas que constituyen los verdaderos sensores de los distintos compuestos en el alimento (Cano et al., 2017). La interacción entre compuestos en el alimento y el receptor molecular adecuado de la NSG desencadenará una serie de eventos en el interior de la neurona de forma tal que la señal llegue al cerebro del insecto para su procesamiento. Sin embargo, no todos los compuestos, sustancias o moléculas del alimento tendrán un receptor sensible a su detección. Aunque los receptores moleculares de muchos compuestos fagoestimulantes y de aquellos con propiedades aversivas o anti-alimentarias de las vinchucas son en su mayoría desconocidos, conocemos al menos algunas moléculas que disparan comportamientos apetitivos o aversivos, como lo descrito en el Capítulo 1. El ATP, la sal NaCl y KCl son fagoestimulantes (Pontes *et al.*, 2017), es decir, favorecen la alimentación. En cambio, sustancias como la cafeína, quinina, salicina, berberina, teofilina producen respuestas anti-alimentarias (Pontes *et al.*, 2014, Muñoz *et al.*, 2020). Es decir, si una vinchuca prueba una solución que contiene alguna de estas moléculas, el insecto no se alimenta. Aún muchas moléculas presentes en abundancia en la sangre de los vertebrados como la glucosa o pequeños

aminoácidos podrían ejercer una función apetitiva de la ingestión. Otras, en cambio, conocidas por su poder aversivo en otros insectos podrían generar respuestas inhibitorias de la alimentación. Ampliar el conocimiento actual sobre moléculas apetitivas y aversivas de las vinchucas reviste importancia, ya que conocer en mayor profundidad este sistema sensorial vinculado a la aceptación o rechazo del alimento podría brindarnos nuevas herramientas de control de estos vectores. En este Capítulo, estudiaremos el comportamiento de *R. prolixus* a compuestos como la glucosa, principal monosacárido de la sangre, y a los aminoácidos leucina y fenilalanina, componentes del plasma de la sangre de vertebrados. Además de su presencia en la sangre, estos compuestos demostraron producir una fuerte preferencia gustativa en la hembra del mosquito *Ae. aegypti* (Ignell *et al.*, 2010). Por otra parte, estudiaremos compuestos aversivos como la quinina (Pontes *et al.*, 2014, Kessler *et al.*, 2014), o de desconocido efecto aversivo en vinchucas, como el extracto de ajo y el metil salicilato. El ajo y el metil salicilato han sido descritos como agonistas de receptores (TRPA1) con función nociceptiva en *D. melanogaster* (Kim *et al.*, 2010).

El objetivo de este capítulo es implementar el programa descrito en el Capítulo 2 en el análisis del comportamiento de alimentación de las vinchucas. De esta manera, clasificamos los distintos compuestos entre apetitivos y aversivos en función de los análisis de EMGs y comprobamos todas las funcionalidades de nuestro *software* MyoRec.

3.2 Materiales y Métodos

3.2.1 Insectos

Se utilizaron ninfas del quinto estadio, según los detalles brindados en el Capítulo 1.

3.2.2 Alimentador Artificial y Sistema de Registro de Electromiografía

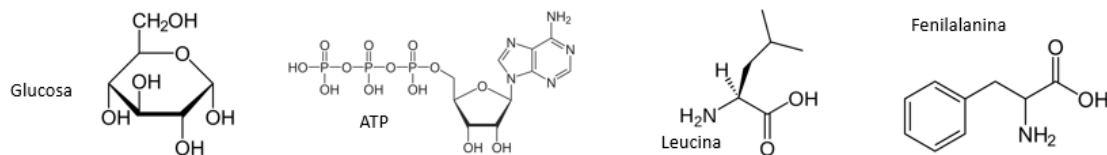
Se utilizó el alimentador artificial y sistema de registro similar al descrito en el Capítulo 1.

3.2.3 Soluciones

Se utilizaron los siguientes compuestos en las siguientes concentraciones: 10 mM glucosa (GLU), 0,33 mM leucina (LEU), 0,33 mM fenilalanina (FEN), 5 mM quinina (QUI), 4 % P/V extracto de ajo (AJO), 1 mM metil salicilato (MES) (**Figura 3.1**). Se utilizó una solución apetitiva control compuesta por: 1 mM ATP y 0,15 M NaCl (CTR), que produce en los insectos un comportamiento altamente apetitivo (Capítulo 1). Cada solución tratamiento fue preparada en esta solución apetitiva control, lo que aseguró una alta motivación alimentaria para los insectos (Friend y Smith, 1971b).

Todas las soluciones se prepararon inmediatamente antes de los experimentos. NaCl (CAS: 7647-14-5) y cafeína (CAS: 58-08-2) se adquirieron en Biopack (Buenos Aires, AR). La glucosa (CAS: 50-99-7) se adquirió de Laboratorios Cicarelli. ATP (CAS: 34369-07-8), leucina (CAS: 61-90-5), fenilalanina (CAS: 63-91-2) y quinina (CAS: 6119-47-7) se compraron en Sigma-Aldrich (St Louis, MO, USA) y el metil salicilato (CAS:119-36-8) en Merck (Buenos Aires, AR).

A

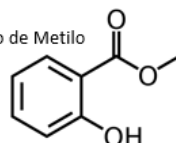


B

Extracto de Ajo



Salicilato de Metilo



Quinina

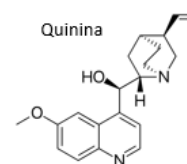


Figura 3.1. Compuestos utilizados en los experimentos. **A** - Compuestos apetitivos. **B** - Compuestos aversivos.

3.2.4 Análisis Estadístico

El número de inserciones (NI), la duración de muestreo (DM), la frecuencia de bombeo (FB), la frecuencia de bombeo en el periodo de contacto (FBC), la amplitud (A), obtenidos a partir de los EMGs, se analizaron estadísticamente utilizando la prueba de Kruskal-Wallis, y se realizaron comparaciones *a posteriori* con la prueba de Dunn. El nivel de significancia fue del 5 % para evaluar la hipótesis nula (H_0 : No existen diferencias entre soluciones tratamiento).

3.3 Resultados

3.3.1 Descripción de los Parámetros y Variables para el Análisis Cuantitativo del Comportamiento

Con la finalidad de describir las diferencias en el comportamiento alimenticio de *R. prolixus* y cuantificar con precisión la información proporcionada por las fases alimenticias observadas, definimos las siguientes variables: **Número de Bombeos (NB)** y **Duración de la Ingestión (DI)**, **Frecuencia de Bombeo (FB)**, **Frecuencia de Bombeo durante el Período de Contacto (FBC)**, **Número de Inserciones (NI)**, **Periodo de Contacto (PC)**, **Duración del Muestreo (DM)**, y la **Amplitud (A)**.

El Número de Bombeos, NB, comprende el total de succiones o contracciones musculares que realiza un individuo durante la FVI, lo cual transcurre durante un período dado por la DI. Ambas variables son utilizadas para calcular la FB, que es el cociente entre el NB y la DI. De la observación del comportamiento alimenticio de los insectos se desprende que en ciertos casos es posible subestimar la FB, debido a que los insectos pueden realizar pausas durante la alimentación a pesar que la ingestión completa no finalizó. Es por ello, que se definió adicionalmente la variable FBC mediante el cociente del parámetro NB y la suma de todos los PC. La diferencia entre la FB y la FBC radica en poder distinguir entre el bombeo efectivo (FBC), de un comportamiento de bombeo donde eventualmente el insecto pueda hacer pausas de alimentación (FB). Si el insecto no realiza pausas de ingestión ambos valores deberían ser iguales.

Por otro lado, se definió la variable NI que permite contabilizar las veces que el insecto inserta sus estiletes en el alimentador artificial. Tanto el evento de inserción como el de

extracción de los estiletes del insecto se caracterizan por un salto de voltaje (que puede ser positivo o negativo con respecto a la línea de base). Este cambio de voltaje se refiere al momento en el cual el circuito eléctrico se cierra o se abre en el EMG.

Por último, se definió la variable A de los picos, dado que la amplitud de cada pico se correlaciona con la fuerza de bombeo, sería esperable encontrar diferencias en esta variable según la solución ofrecida a los insectos.

A continuación se presenta una tabla que muestra las variables definidas.

Tabla 1. Variables definidas para el análisis del comportamiento de alimentación. Se muestra el nombre y el acrónimo.

Nombre	Acrónimo
Periodo de Contacto	PC
Número de Inserciones	NI
Duración del Muestreo	DM
Número de Bombeos*	NB
Duración de la Ingestión*	DI
Frecuencia de Bombeo	FB
Frecuencia de Bombeo en el PC	FBC
Amplitud	A

3.3.2 Comportamiento de Alimentación hacia Sustancias de Diferente Valencia.

Entre los compuestos que categorizamos *a priori* como apetitivos se encuentran el monosacárido glucosa (GLU) y los aminoácidos fenilalanina (FEN) y la leucina (LEU). Entre los compuestos que categorizamos *a priori* como aversivos se encuentran la quinina (QUI), el metil salicilato (MES) y extracto de ajo (AJO).

Utilizando MyoRec se detectaron las distintas fases de alimentación (*i.e.* PC, FM, FVI) en los EMGs para cada insecto tratado con una solución diferente en el alimentador artificial.

Además, el programa admitió el cálculo de las distintas variables para cada grupo experimental, lo que permitió el análisis cuantitativo y comparado que se detalla en la **Figura 3.2**.

En la **Figura 3.2** se muestra el análisis de la respuesta comportamental de los insectos hacia las distintas soluciones.

Por un lado, y de manera global para todas las variables analizadas, es decir NI, DM, FB, FBC, A, no se observaron diferencias estadísticamente significativas entre las soluciones GLU, FEN, LEU y la solución apetitiva control CTR en el comportamiento de alimentación ($p > 0,05$) (**Figura 3.2A-D**). Por otro lado, cuando se compararon las soluciones clasificadas como aversivas, QUI, MES y AJO, contra la solución CTR se observaron diferencias estadísticamente significativas en el comportamiento de alimentación en las siguientes variables: NI, FB, FBC y la A, siendo estas diferencias dependientes del compuesto en cuestión (**Figura 3.2F-I**). Los insectos mostraron un NI mayor hacia el AJO con respecto a la solución CTR. Sin embargo, esta diferencia fue marginalmente significativa ($KW = 7$, $gl = 3$, $p = 0,06$; comparación *a posteriori* de Dunn $p < 0,04$). Además, se observó que los insectos mostraron FB significativamente menores hacia el AJO y la QUI que la solución CTR ($KW = 26$, $gl = 3$, $p < 0,0001$; comparación *a posteriori* de Dunn $p < 0,0001$ y $p = 0,0012$ para AJO y QUI, respectivamente). Si bien no se observó una reducción estadísticamente significativa hacia el MES, sí se observó una tendencia hacia su disminución. Se observó un resultado similar para los mismos compuestos en el caso de la FBC, es decir, que la frecuencia de bombeo durante el periodo de contacto resultó significativamente inferior hacia las soluciones con AJO y QUI con respecto a la solución control ($KW = 21$, $gl = 3$, $p = 0,0001$; comparación *a posteriori* de Dunn $p < 0,0001$ y $p = 0,045$ para AJO y QUI, respectivamente). Las amplitudes de las contracciones de las bombas de los insectos durante la alimentación fueron marginalmente diferentes entre los tratamientos ($KW = 7$, $gl = 3$, $p = 0,07$). Sin embargo, las comparaciones *a posteriori* revelaron diferencias entre la solución CTR y la solución con QUI (comparación *a posteriori* de Dunn $p = 0,041$). Finalmente, no se observaron diferencias entre las soluciones CTR, AJO, QUI y MES en cuanto a la DM ($p > 0,05$). Los resultados aquí presentados mostraron, tanto para las soluciones apetitivas como aversivas, que la DM no sería modulable

por la presencia de los distintos compuestos en la solución (como se observó en el Capítulo 1, **Figura 1.6**).

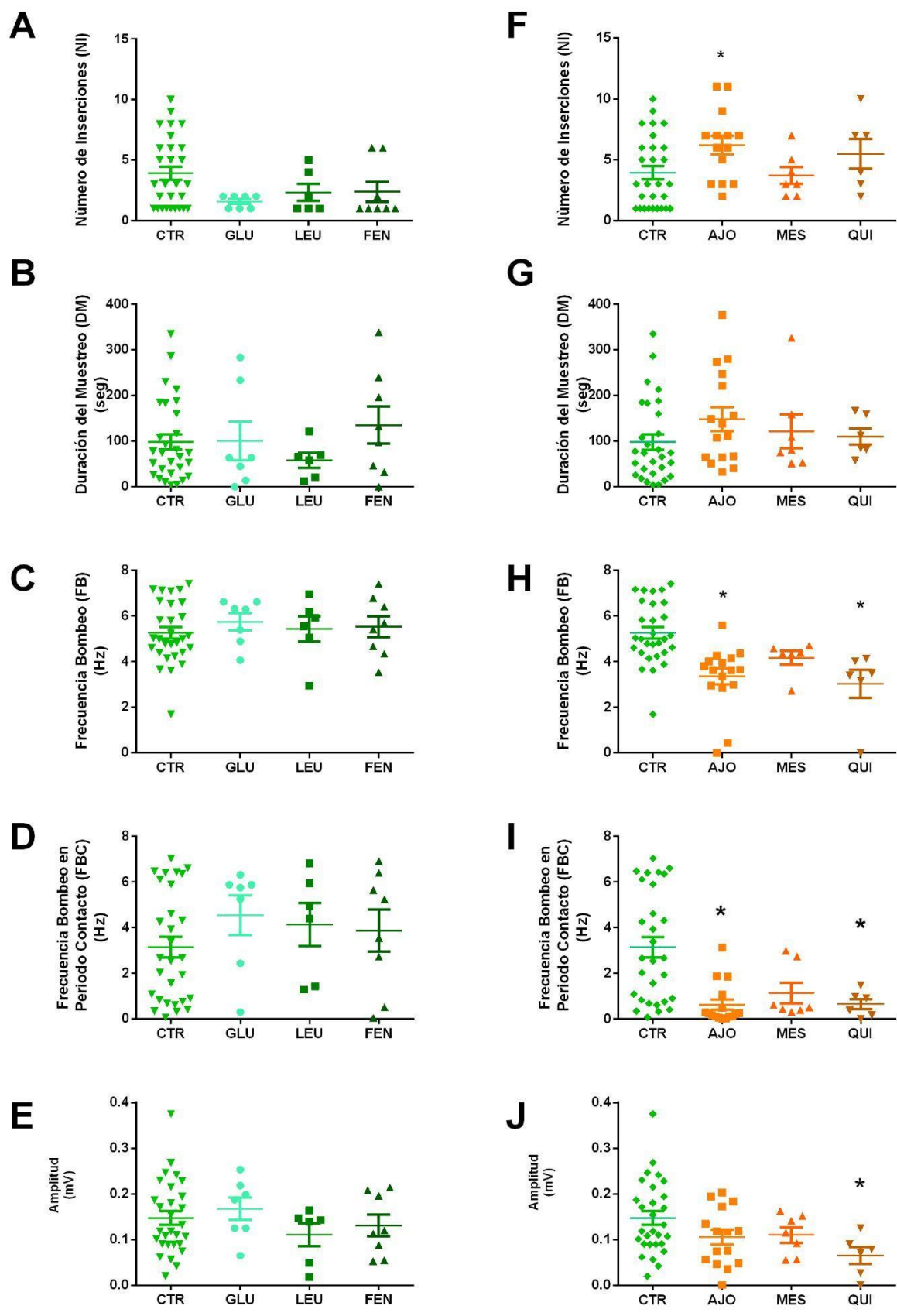


Figura 3.2. Respuesta de alimentación hacia compuestos apetitivos (verde) y aversivos (naranja) respecto a una solución control. Se representa el valor medio ($\pm$ EE) de las siguientes variables en función de las soluciones estímulo ofrecidas: Número de Inserciones (NI), Duración del Muestreo (DM), Frecuencia de Bombeo (FB), Frecuencia de Bombeo durante el Período de Contacto (FBC), Amplitud (A). **A - F** Soluciones apetitivas: glucosa (GLU, N = 7), leucina (LEU, N = 6), fenilalanina (FEN, N = 8). **E - I** Soluciones Aversivas: extracto de ajo (AJO, N = 16), metilsalicilato (MES, N = 7), quinina (QUI, N = 6). En todos los gráficos también se representa la respuesta a la solución apetitiva control (CTR) con el fin de tener una medida de comparación. Los valores correspondientes a cada registro realizado se muestran en cada figura como puntos con distinta forma geométrica. Los asteriscos señalan diferencias significativas con respecto al CTR.

3.4 Discusión

En este Capítulo, evaluamos el comportamiento de alimentación de *R. prolixus* frente a una batería de compuestos cuyos efectos eran parcialmente conocidos o desconocidos, pero que asumimos *a priori* que pertenecían a una de dos categorías posibles: apetitivos o aversivos. Un estímulo apetitivo desencadenará la verdadera ingestión mientras que un aversivo no, lo que implica que la información contenida en la molécula que ingresa al canal alimenticio es detectada en un primer momento por un sistema sensorial específicamente sintonizado a tal fin. Y, posteriormente esa información viaja al cerebro del insecto donde será debidamente procesada de forma tal que el animal podrá tomar una decisión final: ingerir o no ingerir. Sabemos que existen muchos compuestos con propiedades apetitivas, así como antialimentarias o aversivas para insectos hematófagos (Kessler *et al.*, 2014; Ignel *et al.*, 2010; Pontes *et al.*, 2014; 2017; Asparch *et al.*, 2016), sin embargo, todavía no sabemos cómo éstos afectan el comportamiento de los músculos succionadores en la cabeza de dichos insectos. La adaptación del alimentador artificial para obtener los registros de los EMGs de los músculos en combinación con MyoRec para el análisis de los datos nos permitió describir el efecto de compuestos de distinta valencia en el comportamiento de alimentación.

Cuando se estudian los efectos de varios tratamientos en el comportamiento alimenticio de los insectos, es importante contar con información de diferentes variables que permitan observar en detalle los diferentes eventos y mecanismos que subyacen durante el proceso. Una forma simple de observar los efectos en el comportamiento de alimentación de una solución en

los insectos es la medición de su ganancia en peso, es decir pesando los insectos antes y después de la exposición a la solución a probar (Pontes *et al.*, 2014). Esta variable es notablemente contundente para observar diferencias y determinar la categoría de la solución alimenticia (apetitiva o aversiva). Sin embargo, a través de las variables que definimos en esta Tesis de los EMGs, estos registros proveen información adicional que amplifica el número de características observables inherentes al comportamiento alimenticio con las que las soluciones se pueden diferenciar. En el caso de las sustancias catalogadas *a priori* como apetitivas, no se observaron diferencias estadísticamente significativas respecto a la solución control para ninguna de las variables. Esto implica que sus efectos no potenciaron ni synergizaron el efecto fagoestimulante *per se* de la solución de ATP y NaCl, aunque tampoco actuaron en su detrimento o disminución. La solución fagoestimulante por excelencia en *R. prolixus* contiene ATP y NaCl. En promedio, esta solución provoca que los músculos succionadores bombeen a una determinada frecuencia que se traduce en una considerable ganancia de peso, es decir, los insectos pueden ganar entre 10 a 15 veces su peso inicial en 10 minutos que dura el experimento (Pontes *et al.*, 2014). Friend (1965) fue el primero en determinar que la presencia de ATP y solución salina era capaz de provocar la ingestión en el triatomino *R. prolixus*. Más adelante Smith (1970) realizó los primeros registros eléctricos de la bomba cibarial, a través de lo que denominamos EMGs. Smith observó que los registros capturados hacia una solución de ATP y salina eran similares a los producidos por un insecto alimentado con sangre de un hospedador vivo. Desde los trabajos pioneros de Friend y Smith a mediados de los años 60 a la actualidad, se considera al ATP el fagoestimulante más potente para *R. prolixus*.

La glucosa es una molécula de importancia vital para la mayoría de los organismos debido a que es el azúcar más ampliamente utilizado como fuente de energía. Varios estudios han revelado que la ingesta de glucosa por parte de algunos insectos estimula su eficiencia reproductiva y proporciona la energía que los huevos necesitan para su maduración, sin tener que recurrir a las reservas de lípidos (Li *et al.*, 2020; Huang *et al.*, 2015; Thompson, 2003). La glucosa es uno de los compuestos presentes en la sangre que son esenciales para disparar la alimentación en el mosquito *Aedes aegypti* (Friend y Smith, 1977). En *Anopheles gambiae* se han observado receptores gustativos en la labela sensibles a mezclas de fructosa y glucosa (Kessler y

Vlimant, 2015). Es importante notar que la concentración de glucosa utilizada en este trabajo es equivalente a la presente en la sangre humana.

Los aminoácidos, por su parte, también son moléculas esenciales para la vida de todos los organismos. Los aminoácidos constituyen las proteínas y muchos de ellos participan en una variedad de procesos celulares, incluyendo la transducción intracelular. En la naturaleza existen 20 aminoácidos que reciben el nombre de L-aminoácidos canónicos. Muchos de ellos se producen endógenamente en los organismos. Sin embargo, existe un subgrupo de aminoácidos canónicos que solo pueden obtenerse mediante la ingestión de alimento, que se los denomina aminoácidos esenciales, debido a que no se producen de manera endógena. Nueve de estas moléculas son comunes en mamíferos e insectos (histidina, isoleucina, leucina, lisina, metionina, fenilalanina, treonina, triptófano y valina). En experimentos de comportamiento con *Drosophila melanogaster*, algunos de estos compuestos han resultado ser apetitivos, aversivos o neutros, dependiendo de la concentración utilizada (Amrein, 2016). Debido a la naturaleza hematófaga de los triatomíneos, en esta Tesis utilizamos soluciones compuestas por los aminoácidos fenilalanina y leucina en una concentración similar a la existente en la sangre humana. Ambos aminoácidos son esenciales, y aparecen en el plasma de la sangre luego de la digestión y la absorción de las proteínas en una ingesta (Yoshii *et al.*, 2018). La concentración de fenilalanina se encuentra entre 35 y 80 $\mu\text{M/L}$ en un adulto sano (Cleary *et al.*, 2013), mientras que la leucina oscila de 74 a 196 $\mu\text{M/L}$, respectivamente.

Dado el carácter relevante de la glucosa y de 2 aminoácidos esenciales (FEN y LEU) nos preguntamos si estos compuestos podrían tener un efecto positivo en la aceptación del alimento. Sin embargo, los resultados mostraron que ninguno de ellos contribuye adicionalmente a la aceptación de la solución apetitiva de ATP y salina, donde la respuesta es máxima. Es posible, que el sistema gustativo de las vinchucas que se alimentan de un sistema homeostático, es decir, sujeto a pocas variaciones en su composición, esté específicamente sintonizado simplemente a ATP y NaCl, y que no sea sensible a la detección de otros compuestos fagoestimulantes. La GLU, FEN y LEU podrían ser consideradas como neutras, sin un efecto ni positivo ni negativo en la decisión alimenticia, sin un valor sensorial y funcional para *R. prolixus*. Es posible, que las

vinchucas no posean receptores quimiosensoriales sensibles a estos tres compuestos como ocurre en otros insectos (DeBrito Sanchez, 2011; Chen y Dahanukar, 2019; Ganguly *et al.*, 2017; entre otros). Por otra parte, es factible también, que existan otros compuestos de valencia positiva que no se encuentran entre los probados en esta Tesis.

Los resultados fueron más heterogéneos hacia las sustancias clasificadas inicialmente como aversivas. Los efectos observados con la existencia de al menos una diferencia estadísticamente significativa en todas las variables respecto al control, salvo la DM, proporciona evidencia adicional de que estas sustancias son probablemente inhibidoras de la alimentación (**Figura 3.2**). Sin embargo, es evidente que no todas lo son con la misma intensidad. Esto significa que existen diferentes grados de aversión y sustancias como el AJO o la QUI muestran efectos más pronunciados en las variables alimenticias. El MES, por su parte, ejerció una incipiente reducción de las FB y FBC en el comportamiento de los insectos, aunque no fue significativa. Las variables NI, A, y en particular la DM, fueron menos sensibles que la FB y la FBC para diferenciar las soluciones aversivas que probamos respecto a la solución apetitiva control. Sin embargo, la sensibilidad de las variables puede variar con otros tratamientos a pesar de que pertenezcan *a priori* a la misma categoría.

Algunos de los compuestos aversivos utilizados en este trabajo fueron escogidos en base a los efectos registrados tanto en insectos fitófagos como hematófagos. Los compuestos aversivos activan directamente las neuronas sensoriales gustativas (NSGs) o interfieren con la detección de los estímulos apetitivos (French *et al.*, 2015). Sustancias como la cafeína y la QUI ya han sido utilizadas previamente en triatomíneos para evaluar sus efectos en la alimentación, el metabolismo, la supervivencia y en la toma de decisiones (Muñoz *et al.*, 2020). El blanco molecular o receptor de estas sustancias no es conocido en todos los insectos. Sin embargo, y según el conocimiento que tenemos a partir de la mosca de la fruta *D. melanogaster*, muchos de estos compuestos catalogados antropocéntricamente como amargos son detectados por receptores de tipo Gr e Ir (Shrestha y Lee 2023). Muchos de estos compuestos son producidos como metabolitos secundarios de plantas, posiblemente con un rol defensivo contra la herbivoría. En consecuencia, provocan respuestas aversivas en insectos fitófagos que intentan

alimentarse de ellas, aunque también se demostró que pueden generar efectos inhibidores de la alimentación en insectos hematófagos (Pontes *et al.*, 2014, Plata-Rueda *et al.*, 2017, Muñoz *et al.*, 2020). El impacto de estas sustancias amargas en los insectos puede ser diferente. Por ejemplo, en *R. prolixus* la cafeína tiene un efecto antialimentario y tóxico, ya que disminuye la sobrevivencia de los insectos. Por otra parte, la QUI y la teofilina producen también efectos repulsivos sin afectar parámetros fisiológicos en los insectos (Muñoz *et al.*, 2020). Asparch y colaboradores (2016) demostraron mediante un protocolo de condicionamiento no asociativo que *R. prolixus* es capaz de discriminar entre la cafeína de la QUI, esta capacidad podría explicarse gracias a la presencia de diferentes receptores o de diferentes mecanismos de transducción para ambas moléculas en las NSGs. La cafeína y la QUI también disminuyen la alimentación cuando están presentes en soluciones de sacarosa en *Ae. Aegypti* (Ignell *et al.*, 2010), mientras que la QUI reduce la ingestión en *An. gambiae* cuando se le suministra en una solución artificial (Kessler *et al.*, 2014), o adicionada a la sangre (Dennis *et al.*, 2019).

El ajo y sus principales constituyentes han demostrado ser tóxicos y repelentes para varias especies de insectos de importancia agroeconómica, y por lo tanto son utilizados para el control de plagas en depósitos de grano y otros cultivos (Plata-Rueda *et al.*, 2017). El ajo y sus componentes, al igual que muchos aceites esenciales, tiene propiedades ecotoxicológicas favorables, como la baja toxicidad en humanos, pronta degradación y bajo impacto ambiental. Entre sus constituyentes se destacan la alicina y el dialil sulfuro que son agonistas de los receptores TRPA1 (García-Añoveros y Nagata, 2007; Bautista *et al.*, 2005). La activación de dicho receptor tiene un efecto nociceptivo en el insecto que deriva en un estado conocido como hiperalgesia. La presencia de TRPs en el canal alimentario de *R. prolixus* (Ortega Insaurrealde *et al.*, 2023) sugiere que este receptor podría ser el candidato involucrado en la detección del ajo en las sensillas gustativas presentes en la faringe de los insectos, traducándose en la reducción del número de bombeos de los músculos suctores.

El MES es un compuesto ubicuo, frecuentemente producido por plantas que se defienden de la herbivoría de insectos como áfidos y coleópteros, o de ácaros, y puede acarrear efectos olfativos repelentes (Ninkovic *et al.*, 2021). Este compuesto también se encuentra entre las

moléculas agonistas del receptor TRPA1, y podría estar relacionado con la inducción de respuestas nociceptivas en insectos (García-Añoveros y Nagata, 2007). Por esta razón clasificamos esta sustancia *a priori* como aversiva. Sin embargo, el efecto aversivo del MES en el comportamiento de alimentación de *R. prolixus* no fue tan marcado como con respecto al AJO. De todas maneras, sería interesante en el futuro probar el efecto de concentraciones superiores a la ensayada en esta Tesis.

Independientemente de la sustancia aversiva y del receptor molecular involucrado, un sistema sensorial sintonizado en la detección de sustancias, cuya ingestión produce en los insectos condiciones fisiológicas negativas, es ventajoso para prevenir riesgos que le cueste la vida al animal.

Para establecer un patrón sobre la importancia de las variables es necesario contar con un gran número de tratamientos analizados. Los resultados hasta aquí presentados proveen evidencias de que la DM sería la variable menos discriminadora, mientras que la FB y la FBC las de mayor relevancia. Además de proveer información sobre diferentes aspectos del comportamiento alimenticio del insecto, las variables definidas podrían proporcionar información relevante sobre efectos de las sustancias a nivel fisiológico y estimular el postulado de nuevas hipótesis y preguntas. Con el uso de MyoRec, esta información puede obtenerse de forma rápida y precisa, mientras permite el fácil manejo de un gran volumen de datos.

Discusión Final

A continuación se muestran las hipótesis descritas previamente en la introducción y su aceptación o rechazo de acuerdo a los resultados obtenidos:

Hipotesis 1: La decisión de un insecto de alimentarse está controlada por la detección de determinadas moléculas, las cuales desencadenan respuestas apetitivas, *i.e.* aceptación del alimento, o respuestas aversivas, *i.e.* de evitación de la ingestión. Ambos tipos de respuestas producen patrones identificables y característicos de cada condición alimenticia producto de la activación de los músculos involucrados.

Los electromiogramas típicos de cada condición alimenticia son claramente distinguibles con las variables que se definieron definidas. Por lo tanto la hipótesis es verdadera.

Hipótesis 2: Es posible caracterizar la complejidad del comportamiento alimenticio a través de modelos matemáticos.

Diversos modelos matemáticos fueron utilizados para identificar las Fases de Verdadera Ingestión, detectar picos en los electromiogramas y calcular las variables, lo que significa que esta hipótesis también es válida.

Hipótesis 3: Diferentes modelos matemáticos son capaces de detectar patrones en series o señales temporales de cualquier tipo, entonces es posible elaborar un identificador de patrones que reconozca y delimite automáticamente los comportamientos de alimentación.

Se desarrolló el Identificador de Eventos de Verdadera Ingestión (IEVI) que es capaz de identificar de forma automática las Fases de Verdadera Ingestión. A pesar de que los Períodos de Contacto todavía deben establecerse de forma manual, una parte importante de los comportamientos de alimentación es identificada automáticamente por el IEVI, por lo tanto la hipótesis es válida.

El uso de EMGs y la definición de nuevas variables amplía el número de características observables relacionadas al comportamiento alimenticio de los insectos cuando se alimentan de compuestos apetitivos y aversivos. Esto equivale a incrementar los detalles con los que observamos las propiedades de ambos tipos de eventos, acarreado la aparición de nuevas preguntas y perspectivas sobre los fenómenos fisiológicos y de comportamiento que tienen lugar durante el proceso de alimentación. La diferencia en la duración del PC podría deducirse de forma hipotética tras observar que un insecto come más que el otro, o al observar cualitativamente el perfil de los EMGs y discernir que uno tiene más picos y FVIs más largas, aunque menos numerosas. Sin embargo, las hipótesis requieren una comprobación cuantitativa y este es el beneficio que nos brinda la medición del PC: la corroboración empírica de que lo que se intuye es realmente así. Por tanto, la reducción del PC y la DI es un resultado esperado aunque no es el único posible para responder *a priori* la pregunta sobre el por qué un insecto come menos que el otro, aún cuando la explicación se restringe al uso de las variables que definimos. Por ejemplo, el insecto podría comer con la misma duración del PC pero reducir la frecuencia de bombeo de la solución alimenticia, dando como resultado la ingestión de una porción total más pequeña del alimento. También se podría observar este resultado si disminuye únicamente la intensidad (amplitud de los picos de voltaje en los EMGs) con la que bombea, o si se producen ambos efectos al mismo tiempo. Se conocen insectos hematófagos que pueden variar drásticamente la cantidad de líquido ingerido por unidad de tiempo (tasa de ingestión), proporcionando ventajas energéticas y en el tiempo de alimentación según lo exigen las circunstancias (Kikuchi *et al.*, 2018). Este y otros efectos físico-químicos que también afectan la tasa de ingestión, como los cambios de viscosidad del alimento, la presión opuesta producida por el alimento y las paredes cuticulares del abdomen a medida que se llena, el grado de expresión de anticoagulantes y antiinflamatorios en la saliva, el tamaño del vaso sanguíneo, el tamaño de los eritrocitos y su capacidad para deformarse, etc., podrían pasar desapercibidos si solo se observa la duración del PC.

Es indudable la necesidad de la identificación y cuantificación de diversas variables para describir satisfactoriamente el comportamiento alimenticio del animal. Sin embargo, al observar diferentes propiedades de este proceso con las variables que definimos, también es posible

encontrar resultados no esperados o poco intuitivos. Nuestros resultados indican que el número de veces que el insecto atraviesa la membrana del alimentador artificial, es independiente de que la solución alimenticia sea apetitiva o aversiva, aunque en el último caso deben producirse en un tiempo más limitado debido a la reducción del PC. Entonces, sería factible deducir que el NI relativo al PC, y no el NI absoluto, podría ser menor hacia las soluciones apetitivas. Sin embargo, al re-analizar los datos encontramos que sólo la solución de AJO presentó una significativa reducción en el NI/PC con respecto a la solución CTR ($p < 0,05$). El bajo o nulo efecto entre soluciones de ambas categorías de la duración del muestreo (DM) también es un resultado poco intuitivo. Pero nuevamente al recalcular la variable DM y relativizarla al PC encontramos sólo diferencias significativas entre la solución de AJO y la CTR ($p < 0,05$). Por lo tanto, es interesante destacar que estas diferencias sutiles deberán ser ajustadas y tenidas en cuenta en el futuro. Evidentemente la fuerza o poder aversivo de cada compuesto es particular y afectarán el comportamiento de los insectos de manera específica. Es importante entonces que las variables definidas nos permitan caracterizar con más exactitud cuales son los eventos que verdaderamente están ocurriendo entre muchas posibilidades.

Para incrementar los detalles con los que se puede estudiar el comportamiento alimenticio en *R. prolixus* y otras especies de triatominos, e incluso en otros tipos de insectos, no necesariamente hematófagos, se podrían registrar con más detalle los movimientos que realiza el animal y sus partes bucales durante la alimentación, ya sea en un alimentador artificial y/o en la piel de un hospedador. Registrar al mismo tiempo las sensaciones percibidas por el hospedador en caso de que sea humano, y/o utilizar una cámara que permita observar los eventos de forma sincronizada con los observados en los EMGs. Esto podría derivar en la definición de nuevos patrones de onda asociados a nuevos comportamientos y dar lugar a la creación de futuros motores de identificación automática en el IEVI y nuevas variables que permitan cuantificarlos. Por ejemplo, facilitaría el etiquetado de los eventos de inserción y extracción de probóscide para la creación de un identificador automático de PCs.

MyoRec es un programa informático que facilita y reduce considerablemente el tiempo que toma el análisis de EMGs en triatominos, aunque también podría servir para analizar señales

EMGs provenientes de otros insectos hematófagos. A pesar de que aún es necesaria la supervisión del investigador en dichos análisis, MyoRec ofrece una serie de ventajas con respecto a los análisis realizados en forma manual o con el uso de programas no especializados:

1) Proporciona buscadores automáticos y herramientas de selección manual para definir intervalos de tiempo (FVIs), que identifican las fases de alimentación y eventos relevantes registrados en los EMGs.

2) MyoRec utiliza variables equivalentes a las empleadas en trabajos anteriores para contener la información de los registros y agrega algunas nuevas que permiten analizar la dinámica del comportamiento de los insectos durante el tiempo de registro.

3) Posee una aplicación semiautomática (DP) para facilitar la extracción de los datos de los EMGs de forma supervisada, independientemente de la presencia de ruido en la señal, picos deformados y frecuencias de bombeo variables.

4) Incorpora un algoritmo capaz de calcular las variables de forma completamente automática (CD) y con un alto grado de confiabilidad, no solo de un registro electromiográfico, sino de decenas de ellos en tan solo unos minutos, con solo presionar un botón.

5) Almacena la información de los EMGs en archivos estandarizados que contienen todos los datos relacionados a los análisis. Estos archivos pueden ser individuales (información de un solo EMG) o de tratamiento (información de múltiples EMGs). Los datos que contienen sobre las coordenadas de los picos de voltaje y los límites de los eventos en la señal, permiten reproducir y visualizar los análisis en sesiones futuras con el fin de inspeccionar detalles y realizar modificaciones.

6) MyoRec ofrece numerosas opciones gráficas para visualizar información referente a los EMGs. Las visualizaciones de la señal completa de los EMGs, como los espectrogramas y las variaciones temporales de la frecuencia fundamental de la señal, permiten observar aspectos generales del comportamiento de alimentación. Los gráficos en ventanas temporales de la señal permiten observar detalles del rendimiento de los algoritmos de cálculo (e.g. DP y CD), y encontrar valores para los hiperparámetros que los hagan más efectivos. Esto resulta útil

especialmente cuando se tienen que analizar EMGs provenientes de diferentes modelos biológicos que producen perfiles diferentes en los patrones de onda. Las visualizaciones sobre la dinámica temporal de las variables para un EMG o de las medias de múltiples conjuntos de EMGs (tratamientos), caracterizan la tendencia o la evolución en el tiempo de cada variable mientras el insecto se alimenta, lo que deja entrever algunos aspectos del comportamiento que no pueden deducirse mediante los valores acumulativos de las variables. La aplicación para la creación de diagramas de caja de los datos ofrece al usuario un análisis estadístico preliminar que puede ser refinado posteriormente con programas externos más especializados si es necesario.

7) El procesamiento de los EMGs con MyoRec permite etiquetarlos y utilizarlos como nuevas instancias para la transferencia de aprendizaje al modelo del IEVI. Esto significa modificar algunos de los pesos del modelo previamente entrenado para que tome en cuenta los nuevos datos en sus predicciones e incremente su efectividad. También hace posible elaborar copias del modelo del IEVI para entrenarlos desde cero con conjuntos de datos provenientes de diferentes insectos. Esto hace que cada modelo se adapte para funcionar en el IEVI con los EMGs de un insecto particular que se está estudiando, aunque bajo el molde de las fases y eventos que describimos para los triatominos. Es importante destacar que esta característica está en fases finales de desarrollo y será implementada en un futuro próximo en el programa.

8) Posee una interfaz gráfica adaptada a las necesidades del usuario para realizar los análisis de forma eficiente y con las funcionalidades más importantes para el tratamiento y el manejo de los datos a su alcance.

En el futuro es posible mejorar el desempeño del modelo del IEVI realizando otras modificaciones a los conjuntos de datos de entrada del modelo. Por ejemplo, es posible que al aplicar la transformada Wavelet en lugar de la transformada de Fourier y/o establecer imágenes de los perfiles que se generan con una o ambos tipos de transformadas como datos de entrada, mejore la calidad del modelo o su capacidad predictiva. Sin embargo, estas modificaciones probablemente acarreen un mayor tiempo de preprocesamiento de los datos para que se les pueda aplicar el modelo y el costo computacional de entrenamiento se incrementará notablemente.

También es posible expandir el número de variables calculadas para ajustarse a la información que se recolecta de forma estándar en otros insectos (Giordanengo, 2014). De esta manera el programa ganaría un carácter más general a medida que todas o la mayor parte de sus funciones puedan ser usadas con una gama más amplia de artrópodos.

Otra de las tareas que puede realizarse a futuro es el desarrollo del identificador de los períodos de contacto (PCs). Esto haría posible que el IEVI abarcara los dos principales eventos que caracterizan al comportamiento alimenticio (PCs y FVIs) de los triatomíneos, abriendo la posibilidad de que MyoRec pueda realizar los análisis de los EMGs de forma completamente automática y en tiempo real, e.g., paralelamente y en simultáneo a la ejecución de los experimentos. Evidentemente, el desempeño de los modelos involucrados en el reconocimiento y delimitación de los comportamientos tendría que ser lo suficientemente alto y confiable, pero es un resultado que podría conseguirse de forma gradual.

También es posible establecer una mejora en el algoritmo del CD para disminuir el error que comete al estimar la frecuencia de bombeo en FVIs de corta duración. La transformada de Fourier tiene la limitación de que al aplicarse en una ventana temporal con una duración similar a la FVI ofrece una baja resolución en el dominio de frecuencias, lo que dificulta obtener la FB con precisión. Sin embargo, al utilizar la transformada Wavelet en lugar de la de Fourier, con una Wavelet adecuada como la de tipo *Mexican Hat*, todas las frecuencias presentes en la señal pueden ser determinadas con alta precisión para cada instante de tiempo, incluyendo la frecuencia fundamental, que equivale a la FB según la convención que aplicamos.

El software MyoRec se puede descargar gratuitamente desde https://github.com/hecsalms/MyoRec_Dist_Version para sistemas Windows. En ese repositorio se puede encontrar el manual de usuario, el conjunto de datos empleado para el entrenamiento del modelo del IEVI y el conjunto de datos que se usó para evaluar el rendimiento del algoritmo del calculador de datos (CD).

El análisis realizado para sustancias de valencias opuestas cumple con el objetivo que nos propusimos sobre emplear el programa desarrollado *ad hoc* para estudiar el comportamiento de

alimentación de las vinchucas. Hay que notar que hasta donde sabemos no se han realizado estudios con la metodología desarrollada en este trabajo con varias de las sustancias que utilizamos, entre ellas glucosa (GLU), fenilalanina (FEN), leucina (LEU), metil salicilato (MES), y el extracto de ajo (AJO). Los experimentos de ingestión ayudan a comprender los mecanismos fisiológicos y psicoquímicos que están involucrados en los efectos que producen estas moléculas en el comportamiento de alimentación. También contribuyen a la identificación de sustancias repelentes de origen natural y al desarrollo de productos sintéticos basados en ellas. Esto se perfila como una alternativa al uso de insecticidas como estrategia de control vectorial, dado el creciente aumento de casos de resistencia observados en los insectos (Reisenman et al., 2016). Los experimentos realizados en esta tesis están enmarcados en un esfuerzo internacional para desarrollar compuestos repelentes efectivos que sean capaces de impedir el contacto o la picadura de este tipo de animales. Las investigaciones en el desarrollo de sustancias repelentes se han enfocado principalmente en el control de mosquitos, principales insectos vectores de enfermedades. Esto ha permitido la experimentación con miles de químicos que derivaron en la identificación de múltiples repelentes químicos, entre los que destaca el N,N-diethyl-metatoluamide (DEET) (Norris y Coats, 2017). Nuevos compuestos repelentes han resultado ser igualmente efectivos al DEET en triatominos, pero a concentraciones considerablemente menores. Sin embargo, se desconocen los mecanismos de acción que utilizan (Melanie et al., 2020). Las propuestas sobre estrategias de control alternativas en insectos hematófagos pretenden actuar en diferentes modalidades sensoriales (e.g., gusto, olfato, visión). Entre ellas destaca el desarrollo de compuestos repelentes espaciales y de contacto, así como el uso de trampas basadas en la emisión de claves olfativas y visuales para el monitoreo de poblaciones y la eliminación de los insectos (Cardoso et al., 2019; Rojas de Arias et al., 2012). De esta forma, se puede desarrollar una defensa escalonada que involucre a cada modalidad sensorial, de modo que, si falla una de las barreras, el insecto todavía debe atravesar otros obstáculos para llegar al hospedador. Al representar la última frontera de decisión entre el insecto y el hospedador, el sentido del gusto cobra una relevancia especial y esto, sumado al reducido número de trabajos realizados para comprender este mecanismo en los insectos hematófagos, hace que este trabajo sea una valiosa contribución en el desarrollo de las investigaciones asociadas a esta modalidad

sensorial en esta clase de insectos. Aunque el AJO, la CAF y la QUI inhiben la alimentación de los triatomíneos en un grado significativo, es necesario continuar investigando acerca del rol que poseen como posibles repelentes, por ejemplo sobre un hospedador vivo. Es posible que existan otras moléculas distintas a las probadas en esta tesis que produzcan efectos aún más intensos que los observados, ya sea de forma individual o al incorporarse en una mezcla. Al mismo tiempo, las moléculas repelentes deben ser inocuas para el ser humano cuando se apliquen sobre la piel. Esto requiere futuros estudios que incluyen experimentos farmacéuticos pero que pueden derivar en la creación de un producto accesible o de bajo costo, capaz de disminuir la prevalencia de la enfermedad de Chagas, especialmente en las poblaciones más desatendidas.

Bibliografía

- Alzogaray, R. A., Zerba, E. N. (2017). *Rhodnius prolixus* intoxicated. *J Insect Physiol*, 97, 93-113. <https://doi.org/10.1016/j.jinsphys.2016.04.004>.
- Adasme-Carreño, F., Muñoz-Gutiérrez, C., Salinas-Cornejo, J., Ramírez, C. C. (2015). A2EPG: A new software for the analysis of electrical penetration graphs to study plant probing behaviour of hemipteran insects. *Computers and Electronics in Agriculture*, 113, 128–135. <https://doi.org/10.1016/j.compag.2015.02.005>
- Amrein, H. (2016). Mechanism of Taste Perception in *Drosophila*. Chemosensory Transduction: The Detection of Odors, Tastes, and Other Chemostimuli. *Elsevier Inc.* <https://doi.org/10.1016/B978-0-12-801694-7.00014-7>
- Araujo, R. N., Soares, A. C., Gontijo, A.F., Guarneri, A.A., Pereira, M.H., Gontijo, N.F. (2011). Electromyogram of the Cibarial Pump and the Feeding Process in Hematophagous Hemiptera. *Advances in Applied Electromyography*. <https://doi.org/10.5772/10647>
- Araujo, R., Gontijo, N. N. de F., Guarneri, A. A. A., Gontijo, A. A. F., Soares, A. C. A., Pereira, M. M. H., ... Pereira, M. M. H. (2011). Electromyogram of the cibarial pump and the feeding process in hematophagous Hemiptera. (J. Mizrahi, Ed.), *Advances in Applied Electromyography*. InTech.
- Asparch, Y., Pontes, G., Masagué, S., Minoli, S., Barrozo, R. B. (2016). Kissing bugs can generalize and discriminate between different bitter compounds. *Journal of Physiology Paris*, 110(3), 99–106. <https://doi.org/10.1016/j.jphysparis.2016.11.006>
- Backus, E. A., Cervantes, F. A., Guedes, R. N. C., Li, A. Y., Wayadande, A. C. (2019). AC–DC electropenetrography for in-depth studies of feeding and oviposition behaviors. *Annals of the Entomological Society of America*, 112(3), 236–248. <https://doi.org/10.1093/aesa/saz009>
- Backus, E. A., & Mclean, D. L. (1982). *The Sensory Systems and Feeding Behavior of Leafhoppers. 1. The Aster Leafhopper, Macrostelus fascifrons Stal (Hornoptera, Cicadellidae)*. *JOURNAL OF MORPHOLOGY* (Vol. 172).
- Barrozo, R. B., Lazzari, C. R. (2006). Orientation response of haematophagous bugs to CO₂: The effect of the temporal structure of the stimulus. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 192(8), 827–831. <https://doi.org/10.1007/s00359-006-0120-y>
- Barrozo, R. B. (2019). Food recognition in hematophagous insects. *Current Opinion in Insect Science*, 34, 55–60. <https://doi.org/10.1016/j.cois.2019.03.001>

- Barrozo, R. B., Lazzari, C. R. (2004a). Orientation behaviour of the blood-sucking bug *Triatoma infestans* to short-chain fatty acids: Synergistic effect of L-lactic acid and carbon dioxide. *Chemical Senses*, 29(9), 833–841. <https://doi.org/10.1093/chemse/bjh249>
- Barrozo, R. B., Lazzari, C. R. (2004b). The response of the blood-sucking bug *Triatoma infestans* to carbon dioxide and other host odours. *Chemical Senses*, 29(4), 319–329. <https://doi.org/10.1093/chemse/bjh035>
- Barrozo, R. B., Minoli, S. A., Lazzari, C. R. (2004). Circadian rhythm of behavioural responsiveness to carbon dioxide in the blood-sucking bug *Triatoma infestans* (Heteroptera: Reduviidae). *Journal of Insect Physiology*, 50(2–3), 249–254. <https://doi.org/10.1016/j.jinsphys.2004.01.001>
- Barrozo, R. B., Reisenman, C. E., Guerenstein, P., Lazzari, C. R., & Lorenzo, M. G. (2017). An inside look at the sensory biology of triatomines. *Journal of Insect Physiology*, 97, 3–19. <https://doi.org/10.1016/j.jinsphys.2016.11.003>
- Barth, R., (1952). Estudos anatômicos e histológicos sobre a subfamília Triatominae (Hemíptera, Reduviidae). II. parte: Um nôvo órgão sensível das Triatominae. *Mem I Oswaldo Cruz*, 1, 1-4.
- Bautista, D. M., Movahed, P., Hinman, A., Axelsson, H. E., Sterner, O., Högestätt, E. D., Zygmunt, P. M. (2005). Pungent products from garlic activate the sensory ion channel TRPA1. *Proceedings of the National Academy of Sciences of the United States of America*, 102(34), 12248–12252. <https://doi.org/10.1073/pnas.0505356102>
- Bennet-Clark, H. C. (1963a). Negative Pressures Produced in the Pharyngeal Pump of the Blood-Sucking Bug, *Rhodnius Prolixus*. *Journal of Experimental Biology*, 40(1), 223–229. <https://doi.org/10.1242/jeb.40.1.223>
- Bennet-Clark, H. C. (1963b). The Control of Meal Size in the Blood Sucking Bug, *Rhodnius Prolixus*. *Journal of Experimental Biology*, 40(4), 741–750. <https://doi.org/10.1242/jeb.40.4.741>
- Benton, R. (2017). The neurobiology of gustation in insect disease vectors: progress and potential. *Current Opinion in Insect Science*, 20, 19–27. <https://doi.org/10.1016/j.cois.2017.02.003>
- Bodin, A., Barrozo, R. B., Couton, L., Lazzari, C. R. (2008). Temporal modulation and adaptive control of the behavioural response to odours in *Rhodnius prolixus*. *Journal of Insect Physiology*, 54(9), 1343–1348. <https://doi.org/10.1016/j.jinsphys.2008.07.004>
- Cano, A., Pontes, G., Sfara, V., Anfossi, D., Barrozo, R. B. (2017). Nitric oxide contributes to high-salt perception in a blood-sucking insect model. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-15861-0>

- Cardozo, M., Fiad, F. G., Crocco, L. B., Gorla, D. E. (2020). Attraction of *Triatoma infestans* (Klug) to adhesive yeast-baited trap under laboratory conditions. *Int J Trop Insect Sci*, 40, 209-215. <https://doi.org/10.1007/s42690-019-00071-0>
- Castro, J. A., Montalto de Mecca, M., Gómez, M. I. D., Castro, D. G. (2015). Enfermedad de Chagas: Contribuciones del Centro de Investigaciones Toxicológicas. *Acta bioquím. clín. Latinoam*, 49(1). 73-82.
- Cuthbert, R.N., Darriet, F., Chabrierie, O. et al. (2023). Invasive hematophagous arthropods and associated diseases in a changing world. *Parasites Vectors*, 16, 291. <https://doi.org/10.1186/s13071-023-05887-x>
- Chen, Y. C. D., Dahanukar, A. (2020). Recent advances in the genetic basis of taste detection in *Drosophila*. *Cellular and Molecular Life Sciences*, 77(6), 1087–1101. <https://doi.org/10.1007/s00018-019-03320-0>
- Choumet, V., Attout, T., Chartier, L., Khun, H., Sautereau, J., Robbe-Vincent, A., ... Bain, O. (2012). Visualizing non infectious and infectious *Anopheles gambiae* blood feedings in naive and saliva-immunized mice. *PLoS ONE*, 7(12). <https://doi.org/10.1371/journal.pone.0050464>
- de Brito Sanchez, M. G. (2011). Taste perception in honey bees. *Chemical Senses*, 36(8), 675–692. <https://doi.org/10.1093/chemse/bjr040>
- de Oliveira, A. B. B., Alevi, K. C. C., Imperador, C. H. L., Madeira, F. F., Azeredo-Oliveira, M.T.V. (2018). Parasite-Vector Interaction of Chagas Disease: A Mini-Review. *Am J Trop Med Hyg*, 98(3), 653-655. <https://doi.org/10.4269/ajtmh.17-0657>.
- Dell'Arciprete, A., Braunstein, J., Touris, C., Dinardi, G., Llovet, I., Sosa-Estani, S. (2016). Cultural barriers to effective communication between Indigenous communities and health care providers in Northern Argentina: an anthropological contribution to Chagas disease prevention and control. *Int J Equity Health*, 13(6). <https://doi.org/10.1186/1475-9276-13-6>.
- Dennis, E. J., Goldman, O. V., Vossall, L. B. (2019). *Aedes aegypti* mosquitoes use their legs to sense DEET on contact. *Current Biology*, 29(9), 1551-1556. <https://doi.org/10.1016/j.cub.2019.04.004>
- Dukas, R. (2008). Evolutionary biology of insect learning. *Annual Review of Entomology*, 53, 145–160. <https://doi.org/10.1146/annurev.ento.53.103106.093343>
- Ebert, T. A., Backus, E. A., Cid, M., Fereres, A., Rogers, M. E. (2015). A new SAS program for behavioral analysis of electrical penetration graph data. *Computers and Electronics in Agriculture*, 116, 80–87. <https://doi.org/10.1016/j.compag.2015.06.011>

- Falibene, A., & Josens, R. (2008). Nectar intake rate is modulated by changes in sucking pump activity according to colony starvation in carpenter ants. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 194(5), 491–500. <https://doi.org/10.1007/s00359-008-0323-5>
- Ferreira, R. A., Lazzari, C. R., Lorenzo, M. G., Pereira, M. H. (2007). Do haematophagous bugs assess skin surface temperature to detect blood vessels? *PLoS ONE*, 2(9), 1–6. <https://doi.org/10.1371/journal.pone.0000932>
- Flores, G. B., & Lazzari, C. R. (1996). The role of the antennae in *Triatoma infestans*: Orientation towards thermal sources. *Journal of Insect Physiology*, 42(5), 433–440. [https://doi.org/10.1016/0022-1910\(95\)00137-9](https://doi.org/10.1016/0022-1910(95)00137-9)
- French, A., Agha, M. A., Mitra, A., Yanagawa, A., Sellier, M. J., Marion-Poll, F. (2015). Drosophila bitter taste(s). *Frontiers in Integrative Neuroscience*. <https://doi.org/10.3389/fnint.2015.00058>
- Friend, W. G. (1965). The gorging response in *Rhodnius prolixus*. *Canadian Journal of Zoology*, 43(60), 125–132. <https://doi.org/10.1139/z65-010>
- Friend, W. G., & Smith, J. J. (1977). Factors affecting feeding by bloodsucking insects. *Annual Review of Entomology*, 22, 309–331. <https://doi.org/10.1146/annurev.en.22.010177.001521>
- Friend, W. G., Smith, J. J. B. (1971a). Feeding in *Rhodnius prolixus*: Potencies of nucleoside phosphates in initiating gorging. *Journal of Insect Physiology*, 17(7), 1315–1320. [https://doi.org/10.1016/0022-1910\(71\)90196-X](https://doi.org/10.1016/0022-1910(71)90196-X)
- Friend, W. G., Smith, J. J. B. (1971b). Mouthpart activity and salivation, and their correlation with changes of electrical resistance. *J. Insect Physiol*, 17.
- Friend, W. G., & Smith, J. J. B. (1975). Feeding in *Rhodnius prolixus*: Increasing sensitivity to ATP during prolonged food deprivation. *Journal of Insect Physiology*, 21(5), 1081–1084. [https://doi.org/10.1016/0022-1910\(75\)90119-5](https://doi.org/10.1016/0022-1910(75)90119-5)
- García-Añoveros, J., Nagata, K. (2007). TRPA1. *Handbook of Experimental Pharmacology*, 347–362. [doi:10.1007/978-3-540-34891-7_21](https://doi.org/10.1007/978-3-540-34891-7_21)
- Galun, R., Margalit, J. (1969). Adenine nucleotides as feeding stimulants of the tsetse fly *Glossina austeni* Newst. *Nature*, 222, 583–584.
- Galun, R., Kabayo P.J. (1988). Gorging response of *Glossina palpalis* to ATP analogues. *Nature*, 13, 149–423.

- Ganguly, A., Pang, L., Duong, V. K., Lee, A., Schoniger, H., Varady, E., Dahanukar, A. (2017). A Molecular and Cellular Context-Dependent Role for Ir76b in Detection of Amino Acid Taste. *Cell Reports*, 18(3), 737–750. <https://doi.org/10.1016/j.celrep.2016.12.071>
- Giordanengo, P. (2014). EPG-Calc: A PHP-based script to calculate electrical penetration graph (EPG) parameters. *Arthropod-Plant Interactions*, 8(2), 163–169. <https://doi.org/10.1007/s11829-014-9298-z>
- Griffiths, R. B., Gordon, R. M. (1952). An apparatus which enables the process of feeding by mosquitoes to be observed in the tissues of a live rodent; together with an account of the ejection of saliva and its significance in malaria. *Annals of Tropical Medicine and Parasitology*, 46(4), 311–319. <https://doi.org/10.1080/00034983.1952.11685536>
- Guarneri, A. A., Diotaiuti, L., Gontijo, N. F., Gontijo, A. F., Pereira, M. H. (2000). Comparison of feeding behaviour of *Triatoma infestans*, *Triatoma brasiliensis* and *Triatoma pseudomaculata* in different hosts by electronic monitoring of the cibarial pump. *Journal of Insect Physiology*, 46(7), 1121–1127. [https://doi.org/10.1016/S0022-1910\(99\)00222-X](https://doi.org/10.1016/S0022-1910(99)00222-X)
- Guerenstein, P. G., Guerin, P. M. (2001). Olfactory and behavioural responses of the blood-sucking bug *Triatoma infestans* to odours of vertebrate hosts. *Journal of Experimental Biology*, 204(3), 585–597. <https://doi.org/10.1242/jeb.204.3.585>
- Guerrieri, E., Digilio, M. C. (2008). Aphid-plant interactions: A review. *Journal of Plant Interactions*, 3(4), 223–232. <https://doi.org/10.1080/17429140802567173>
- Gorla, D., Hashimoto, K. (2017). Control strategies against Triatominae. In: Telleria, J., Tibayrenc, M. American Trypanosomiasis, Chagas Disease: One Hundred Years of Research. Second Edition. 223–242.
- Guhl, F., Pinto, N., Aguilera, G. (2009). Sylvatic triatominae: a new challenge in vector control transmission. *Mem. Inst. Oswaldo Cruz* 104, 71–75.
- Glendinning, J.J.I., Domdom, S., Long, E., 2001. Selective adaptation to noxious foods by a herbivorous insect. *J. Exp. Biology*, 204, 3355–3367.
- Glendinning, J.I., 2007. How do predators cope with chemically defended foods? *Biol. Bull.* 213, 252–266. <https://doi.org/213/3/252>
- Huang, F. X., Jiang, S., Ren, X. Y., An, T., Chen, H. Y., Zhang, L. S. (2015). Differential expression of proteins associated with lipid metabolism in diapausing *Aphidius gifuensis*. *Chin. J. Biol. Control*, 31, 811–820. <https://doi.org/10.16409/j.cnki.2095-039x.2015.06.001>

- Ignell, R., Okawa, S., Englund, J. E., Hill, S. R. (2010). Assessment of diet choice by the yellow fever mosquito *Aedes aegypti*. *Physiological Entomology*, 35(3), 274–286. <https://doi.org/10.1111/j.1365-3032.2010.00740.x>
- Jové, V., Gong, Z., Hol, F. J. H., Zhao, Z., Sorrells, T. R., Carroll, T. S., ... Vosshall, L. B. (2020). Sensory Discrimination of Blood and Floral Nectar by *Aedes aegypti* Mosquitoes. *Neuron*, 108(6), 1163–1180.e12. <https://doi.org/10.1016/j.neuron.2020.09.019>
- James G., Witten D., Hastie T., Tibshirani R. (2021). An Introduction to Statistical Learning: with Applications in R. *Springer Science+Business Media, LLC*, New York. DOI 10.1007/978-1-4614-7138-7 1
- Kanzaki, R. (1996). Behavioral and neural basis of instinctive behavior in insects: Odor-source searching strategies without memory and learning. *Robotics and Autonomous Systems*, 18(1–2), 33–43. [https://doi.org/10.1016/0921-8890\(95\)00081-X](https://doi.org/10.1016/0921-8890(95)00081-X)
- Kasetty, S., Stafford, C., Walker, G. P., Wang, X., & Keogh, E. (2008). Real-time classification of streaming sensor data. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 1, 149–156. <https://doi.org/10.1109/ICTAI.2008.143>
- Kessler, S., Vlimant, M., & Guerin, P. M. (2013). The sugar meal of the african malaria mosquito *Anopheles gambiae* and how deterrent compounds interfere with it: A behavioural and neurophysiological study. *Journal of Experimental Biology*, 216(7), 1292–1306. <https://doi.org/10.1242/jeb.076588>
- Kessler, S., González, J., Vlimant, M., Glauser, G., & Guerin, P. M. (2014). Quinine and artesunate inhibit feeding in the African malaria mosquito *Anopheles gambiae*: The role of gustatory organs within the mouthparts. *Physiological Entomology*, 39(2), 172–182. <https://doi.org/10.1111/phen.12061>
- Kessler, S., Vlimant, M., & Guerin, P. M. (2015). Sugar-sensitive neurone responses and sugar feeding preferences influence lifespan and biting behaviours of the Afrotropical malaria mosquito, *Anopheles gambiae*. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 201(3), 317–329. <https://doi.org/10.1007/s00359-015-0978-7>
- Khachaturyan, A. G., & Shatalov, G. A. (1971). Elastic Energy of Heterophase Systems of Lamellar Inclusions. *Physics of Metals and Metallography*, 31(6), 1–5. [https://doi.org/10.1016/0010-4361\(73\)90803-3](https://doi.org/10.1016/0010-4361(73)90803-3)
- Kikuchi, K., Stremler, M. A., Chatterjee, S., Lee, W. K., Mochizuki, O., Socha, J. J. (2018). Burst mode pumping: A new mechanism of drinking in mosquitoes. *Scientific Reports*, 8(1), 1–15. <https://doi.org/10.1038/s41598-018-22866-w>

- Kim, S. H., Lee, Y., Akitake, B., Woodward, O. M., Guggino, W. B., Montella, C. (2010). Drosophila TRPA1 channel mediates chemical avoidance in gustatory receptor neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 107(18), 8440–8445. <https://doi.org/10.1073/pnas.1001425107>
- King, B.H., Gunathunga, P.B. (2023). Gustation in insects: taste qualities and types of evidence used to show taste function of specific body parts. *J Insect Sci*, 23(2), 11. <https://doi.org/10.1093/jisesa/iead018>.
- Koul, O. (2005). Insect Antifeedant. *CRC Press LLC*.
- Klowden, M.J. (2007). Physiological Systems in Insects. *Elsevier Inc*. 2007
- Lahondère, C., Insausti, T. C., Paim, R. M. M., Luan, X., Belev, G., Pereira, M. H., ... Lazzari, C. R. (2017). Countercurrent heat exchange and thermoregulation during blood-feeding in kissing bugs. *ELife*, 6, 1–19. <https://doi.org/10.7554/eLife.26107>
- Lavoipierre, M. M. J. (1967). Feeding Mechanism of Haematopinus suis, on the Transilluminated Mouse Ear. *Experimental Parasitology*, 20.
- Lavoipierre (1965). Feeding Mechanism of Blood-sucking Arthropods. *Nature*, 218.
- Lazzari, C. R. (2009). Orientation Towards Hosts in Haematophagous Insects. *An Integrative Perspective. Advances in Insect Physiology*, 37. [https://doi.org/10.1016/S0065-2806\(09\)37001-0](https://doi.org/10.1016/S0065-2806(09)37001-0)
- Li, A. Y., Cook, S. C., Sonenshine, D. E., Posada-Florez, F., Noble, N. I. I., Mowery, J., ... Bauchan, G. R. (2019). Insights into the feeding behaviors and biomechanics of Varroa destructor mites on honey bee pupae using electropenetrography and histology. *Journal of Insect Physiology*, 119, 103950. <https://doi.org/10.1016/j.jinsphys.2019.103950>
- Liang, T., & Brinkman, B. A. W. (2022). Evolution of innate behavioral strategies through competitive population dynamics. *PLoS Computational Biology*, 18. <https://doi.org/10.1371/journal.pcbi.1009934>
- Liman, E. R., Zhang, Y. V., & Montell, C. (2014). Peripheral coding of taste. *Neuron*, 81(5), 984–1000. <https://doi.org/10.1016/j.neuron.2014.02.022>
- Lois-Milevicich, J., Schilman, P. E., & Josens, R. (2021). Viscosity as a key factor in decision making of nectar feeding ants. *Journal of Insect Physiology*, 128, 104164. <https://doi.org/10.1016/j.jinsphys.2020.104164>
- Lehane, M.J., (2005). The biology of blood-sucking in insects. Cambridge University Press.

- Lukenge, M., Ignell, R., Hill, S.R. (2022). Phagostimulants drive the acceptance of a blood meal in disease vectors. In: Sensory ecology of disease vectors. Chapter 17, 469-488, Brill Academic Pub.
- Mans, B. J. (2011). Evolution of vertebrate hemostatic and inflammatory control mechanisms in blood-feeding arthropods. *Journal of Innate Immunity*, 3(1), 41–51. <https://doi.org/10.1159/000321599>
- Marquardt W.C. (2005). Biology of Disease Vectors. *New York: Elsevier Academic Press*.
- McLean, D. L., Kinsey, M. G. (1964). A Technique for Electronically Recording Aphid Feeding and Salivation. *Nature* 202, 1358-9.
- McLean, D. L., Kinsey, M. G. (1967). Probing Behavior of the Pea Aphid, *Acyrtosiphon pisum*. I. Definitive Correlation of Electronically Recorded Waveforms with Aphid Probing Activities 1. *Nature* 2, 400-5.
- Mesquita, R. D., Vionette-Amaral, R. J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., Oliveira, P. L. (2015). Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proceedings of the National Academy of Sciences of the United States of America*, 112(48), 14936–14941. <https://doi.org/10.1073/pnas.1506226112>
- Muñoz, I. J., Schilman, P. E., Barrozo, R. B. (2020). Impact of alkaloids in food consumption, metabolism and survival in a blood-sucking insect. *Scientific Reports*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-65932-y>
- Nava-Doctor, J.E., Sandoval-Ruiz, C.A, Fernández-Crispín, A. (2021). Knowledge, attitudes, and practices regarding vector-borne diseases in central Mexico. *J Ethnobiology Ethnomedicine* 17(45). <https://doi.org/10.1186/s13002-021-00471-y>
- Norris, E.J., Coats, J.R. (2017). Current and future repellent technologies: the potential of spatial repellents and their place in mosquito-borne disease control. *Int J Environ*, 14(2), 124. <https://doi.org/10.3390/ijerph14020124>.
- Ning, X., Selesnick, I. W., & Duval, L. (2014). Chromatogram baseline estimation and denoising using sparsity (BEADS). *Chemometrics and Intelligent Laboratory Systems*, 139, 156–167. <https://doi.org/10.1016/j.chemolab.2014.09.014>
- Ninkovic, V., Glinwood, R., Ünlü, A. G., Ganji, S., Unelius, C. R. (2021). Effects of Methyl Salicylate on Host Plant Acceptance and Feeding by the Aphid *Rhopalosiphum padi*. *Frontiers in Plant Science*, 12, 1–14. <https://doi.org/10.3389/fpls.2021.710268>

- Nunes-da-Fonseca, R., Berni, M., Tobias-Santos, V., Pane, A., Araujo, H.M. (2017). *Rhodnius prolixus*: From classical physiology to modern developmental biology. *Genesis*, 55. <https://doi.org/10.1002/dvg.22995>
- O'Haver T. (2023). Pragmatic introduction to signal processing. *University of Maryland, USA*.
- Ortega-Insaurralde, I., Barrozo, R. B. (2021). The closer the better: Sensory tools and host-association in blood-sucking insects. *Journal of Insect Physiology*, 136. <https://doi.org/10.1016/j.jinsphys.2021.104346>.
- Pérez-Molina, J.A, Crepillo-Andújar, C., Bosch-Nicolau, P., Molina, J. (2021). Trypanocidal treatment of Chagas disease. *Enfermedades Infecciosas y Microbiología Clínica*, 39(9), 458-470. <https://doi.org/10.1016/j.eimc.2020.04.011>
- Petersen, L.R., Jamieson, D.J., Powers, A.M., Honein, M.A. (2016). Zika virus. *J. Med*, 374(16), 1552–1563.
- Pontes, G., Minoli, S., Insaurralde, I. O., De Brito Sanchez, M. G., Barrozo, R. B. (2014). Bitter stimuli modulate the feeding decision of a blood-sucking insect via two sensory inputs. *Journal of Experimental Biology*, 217(20), 3708–3717. <https://doi.org/10.1242/jeb.107722>
- Pontes, G., Pereira, M. H., Barrozo, R. B. (2017). Salt controls feeding decisions in a blood-sucking insect. *Journal of Insect Physiology*, 98, 93–100. <https://doi.org/10.1016/j.jinsphys.2016.12.002>
- Pontes, G., Latorre-Estivalis, J. M., Gutiérrez, M. L., Cano, A., Berón de Astrada, M., Lorenzo, M. G., Barrozo, R. B. (2022). Molecular and functional basis of high-salt avoidance in a blood-sucking insect. *IScience*, 25(7). <https://doi.org/10.1016/j.isci.2022.104502>
- Plata-Rueda, A., Martínez, L. C., Santos, M. H. Dos, Fernandes, F. L., Wilcken, C. F., Soares, M. A., Zanuncio, J. C. (2017). Insecticidal activity of garlic essential oil and their constituents against the mealworm beetle, *Tenebrio molitor* Linnaeus (Coleoptera: Tenebrionidae). *Scientific Reports*, 7, 1–11. <https://doi.org/10.1038/srep46406>
- Rahbr, Y., Febvay, G., Delobel, B., Bonnot, G., (1995). Amino acids and proteins as cues in Aphid-Plant interactions. In: Walker, G. P., Backus, E. A. Recent Research Findings and Experimental Techniques on Homopteran Feeding Behavior, Thomas Say. Publications in Entomology, USA, in press.
- Ramzai J. (2019). Simple guide for ensemble learning methods <https://towardsdatascience.com/simple-guide-for-ensemble-learning-methods-d87cc68705a2>.

- Rojas de Arias, A., Abad-Franch, F., Acosta, N., López, E., González, N., et al. (2012) Post-Control Surveillance of *Triatoma infestans* and *Triatoma sordida* with Chemically-Baited Sticky Traps. *PLOS Neglected Tropical Diseases*, 6(9), e1822. <https://doi.org/10.1371/journal.pntd.0001822>
- Reisenman, C. E., Hong, L., Guerenstein, P. G. (2016). Neuroethology of Olfactory-Guided Behavior and Its Potential Application in the Control of Harmful Insects. *Frontiers in Physiology*, 7. <https://doi.org/10.3389/fphys.2016.00271>
- Ribeiro J.M. (1995). Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect Agents Dis*, 4, 143–152.
- Ribeiro, J. M. C., Francischetti, Ivo. M. B. (2003). Role of Arthropod Saliva in Blood Feeding: Sialome and Post-Sialome Perspectives. *Annual Review of Entomology*. <https://doi.org/10.1146/annurev.ento.48.060402.102812>
- Roberto, M., Sant'anna, V., Diotaiuti, L., De Figueiredo Gontijo, A., De Figueiredo Gontijo, N., & Pereira, M. H. (2001). Feeding behaviour of morphologically similar *Rhodnius* species: influence of mechanical characteristics and salivary function. *Journal of Insect Physiology*, 47. Retrieved from www.elsevier.com/locate/jinsphys
- Rodríguez Ladino, A. C. (2021). Efecto de la alimentación hematófaga y otros comportamientos alimenticios en la eficacia biológica en Triatomíneos *Reduviidae Triatominae*. Una revisión sistemática. Retrieved from <https://ciencia.lasalle.edu.co/biologia/111>
- Romero, A., Chal, C. (2014). Blood constituents as phagostimulants for the bed bug *Cimex lectularius* L. *Journal of Experimental Biology*, 217(4), 552–557. <https://doi.org/10.1242/jeb.096727>
- Sandoval, C. M., Medone, P., Nieves, E. E., Jaimes, D. A., Ortiz, N., & Rabinovich, J. E. (2013). Demographic fitness of *Belminus ferroae* (Hemiptera: Triatominae) on three different hosts under laboratory conditions. *Memorias Do Instituto Oswaldo Cruz*, 108(7), 854–864. <https://doi.org/10.1590/0074-0276130211>
- Sant'Anna, M. R. V., Soares, A. C., Araujo, R. N., Gontijo, N. F., & Pereira, M. H. (2017). Triatomines (Hemiptera, Reduviidae) blood intake: Physical constraints and biological adaptations. *Journal of Insect Physiology*, 97, 20–26. <https://doi.org/10.1016/j.jinsphys.2016.08.004>
- Sarah French, A., Sellier, M. J., Agha Moutaz, A., Guigue, A., Chabaud, M. A., Reeb, P. D., Marion-Poll, F. É. R. (2015). Dual mechanism for bitter avoidance in drosophila. *Journal of Neuroscience*, 35(9), 3990–4004. <https://doi.org/10.1523/JNEUROSCI.1312-14.2015>
- Sarria, E., Cid, M., Garzo, E., & Fereres, A. (2009). Excel Workbook for automatic parameter calculation of EPG data. *Computers and Electronics in Agriculture*, 67(1–2), 35–42. <https://doi.org/10.1016/j.compag.2009.02.006>

- Severson, D.W., Behura, S.K. (2016). Genome investigations of vector competence in *Aedes aegypti* to inform novel arbovirus disease control approaches. *Insects*, 7(4), 58.
- Soares, A. C., Carvalho-Tavares, J., Gontijo, N. de F., dos Santos, V. C., Teixeira, M. M., & Pereira, M. H. (2006). Salivation pattern of *Rhodnius prolixus* (Reduviidae; Triatominae) in mouse skin. *Journal of Insect Physiology*, 52(5), 468–472. <https://doi.org/10.1016/j.jinsphys.2006.01.003>
- Soares, A. C., Araújo, R. N., Carvalho-Tavares, J., Gontijo, N. de F., & Pereira, M. H. (2014). Intravital microscopy and image analysis of *Rhodnius prolixus* (Hemiptera: Reduviidae) hematophagy: The challenge of blood intake from mouse skin. *Parasitology International*, 63(1), 229–236. <https://doi.org/10.1016/j.parint.2013.07.001>
- Schilman, P. E. (2017). Metabolism and gas exchange patterns in *Rhodnius prolixus*. *Journal of Insect Physiology*, 97, 38–44. <https://doi.org/10.1016/j.jinsphys.2016.08.002>
- Schofield C.J., Diotaiuti L., Dujardin J.P. (1999). The process of domestication in Triatominae. *Mem Inst Oswaldo Cruz*, 94(1), 375–378. <https://doi.org/10.1590/S0074-02761999000700073>.
- Shrestha B, Lee Y. (2023). Molecular sensors in the taste system of *Drosophila*. *Genes and Genomics*, 45, 693-707. <https://doi.org/10.1007/s13258-023-01370-0>
- Smith, J. J. B., & Friend, W. G. (1970). Feeding in *Rhodnius prolixus*: Responses to artificial diets as revealed by changes in electrical resistance. *Journal of Insect Physiology*, 16(9), 1709–1720. [https://doi.org/10.1016/0022-1910\(70\)90270-2](https://doi.org/10.1016/0022-1910(70)90270-2)
- Smith, B. Y. J. J. B. (1979). Effect of diet viscosity on the operation of the pharyngeal pump in the blood-feeding bug *Rhodnius prolixus*. *J Exp Biol*, (82), 93-104. <https://doi.org/10.1242/jeb.82.1.93>. PMID: 11799696.
- Snodgrass, R. E. (1935). The Abdominal Mechanisms of a Grasshopper. *Smithsonian Miscellaneous Collections*, 94.
- Tjallingii, W. E. (1988). Electrical recording of stylet penetration activities. In: A. K. Minks & P. Aphids. Their Biology, Natural Enemies and Control, 2B, 95-108.
- Thompson, S.N. (2003). Trehalose : The insect ‘blood’ sugar. *Adv Insect Physiol*, 31, 205-285.
- Van Giessen, W. A., & Jackson, D. M. (1998). Rapid analysis of electronically monitored homopteran feeding behavior. *Annals of the Entomological Society of America*, 91(1), 145–154. <https://doi.org/10.1093/aesa/91.1.145>
- Wayadande, A. C., Backus, E. A., Noden, B. H., Ebert, T., & Hillyer, J. (2020). Waveforms from Stylet Probing of the Mosquito *Aedes aegypti* (Diptera: Culicidae) Measured by AC-DC

Electropenetrography. *Journal of Medical Entomology*, 57(2), 353–368.
<https://doi.org/10.1093/jme/tjz188>

Wigglesworth, V., Gillett, J., 1934. The function of the antennae in *Rhodnius prolixus* (Hemiptera) and the mechanism of orientation to the host. *J Exp Biol*, 11, 120-139.

Wilson, A.L., Courtenay, O., Kelly-Hope, L.A., Scott, T.W., Takken, W., Torr, S.J., Lindsay, S.W. (2020). The importance of vector control for the control and elimination of vector-borne diseases. *PLoS Negl Trop Dis*, 14(1). <https://doi.org/10.1371/journal.pntd.0007831>.

Willett, D. S., George, J., Willett, N. S., Stelinski, L. L., Lapointe, S. L. (2016). Machine Learning for Characterization of Insect Vector Feeding. *PLoS Computational Biology*, 12(11), 1–14.
<https://doi.org/10.1371/journal.pcbi.1005158>

World Health Organization (2017). Global Vector Control Response 2017-2030.

World Health Organization (2020). Vector-borne Diseases.

World Health Organization (2023a). Trypanosomiasis, human African (sleeping sickness).

World Health Organization (2023b). Chagas disease (also known as American trypanosomiasis).

Yoshii, N., Sato, K., Ogasawara, R., Nishimura, Y., Shinohara, Y., Fujita, S. (2018). Effect of mixed meal and leucine intake on plasma amino acid concentrations in young men. *Nutrients*, 10(10), 1–13. <https://doi.org/10.3390/nu10101543>

Apéndice A

A continuación se presenta una breve descripción de los tipos de modelos comparados en el capítulo 2, con el objetivo de que el lector comprenda los fundamentos de funcionamiento y cómo se produce la salida a partir de los datos de entrada.

A.1 Regresión Logística

Debido a que nos proponemos resolver un problema de clasificación, el uso de regresiones lineales no resulta adecuado. En su lugar, optamos por la aplicación de la regresión logística para tratar de clasificar los diferentes segmentos de la señal en los EMGs. Por tanto, el objetivo del algoritmo es modelar la probabilidad de que, dado un segmento de la señal, pertenezca a un comportamiento o no. Es decir, la variable respuesta es discreta binomial. Esto puede escribirse de esta forma:

$$p(X) = P(Y = 1|X) = E(Y = 1|X = x) \quad (1)$$

donde X es el segmento de la señal, Y es la variable binomial, y la probabilidad condicional P equivale a la esperanza E si existe un número suficientemente grande de datos. Una forma de modelar P sería mediante un modelo lineal:

$$p(X) = \beta_0 + \beta_1 X \quad (2)$$

Esto se puede observar aplicando este modelo a un conjunto de datos que simula si un individuo se encuentra en *Incumplimiento* o no respecto a una entidad financiera, de acuerdo a sus ingresos anuales y el saldo mensual de sus tarjetas de crédito (**Figura A.1**). La variable *Incumplimiento* es binaria (0/1) y se pretende predecir el estado de cada individuo con el modelo mencionado. El conjunto de datos cuenta con 10000 observaciones y las personas que incumplieron en un mes determinado se muestran en naranja en la **Figura A.1A**, mientras que los que no lo hicieron en azul. La tasa general de morosidad es de alrededor del 3 %, por lo que se trazó sólo una fracción de los individuos que no incumplieron. Aparentemente las personas que incumplen suelen tener un saldo mayor en sus tarjetas de crédito (**Figura A1.B**), sin embargo, los individuos morosos no

pueden diferenciarse si se tienen en cuenta solamente los ingresos (**Figura A1.C**). Por lo tanto, para nuestro ejemplo no consideraremos la variable *Ingresos* y trataremos de predecir con nuestro modelo el *Incumplimiento* Y para cualquier valor, dado el *Saldo* X . No obstante, como Y no es cuantitativo, el modelo de regresión lineal simple no es una buena elección. Aunque X sea positiva, se producen valores mayores a 1 si es lo suficientemente grande, y valores negativos cuando X es pequeña (**Figura A.2A**). Hay que tener en cuenta que se podría predecir el *Incumplimiento* si en caso de que $p(\text{Saldo} > 0,5)$, aunque si una compañía desea ser más sensible al determinar los individuos que están en riesgo de incumplir, podría fijar un umbral como $p(\text{Saldo} > 0,1)$.

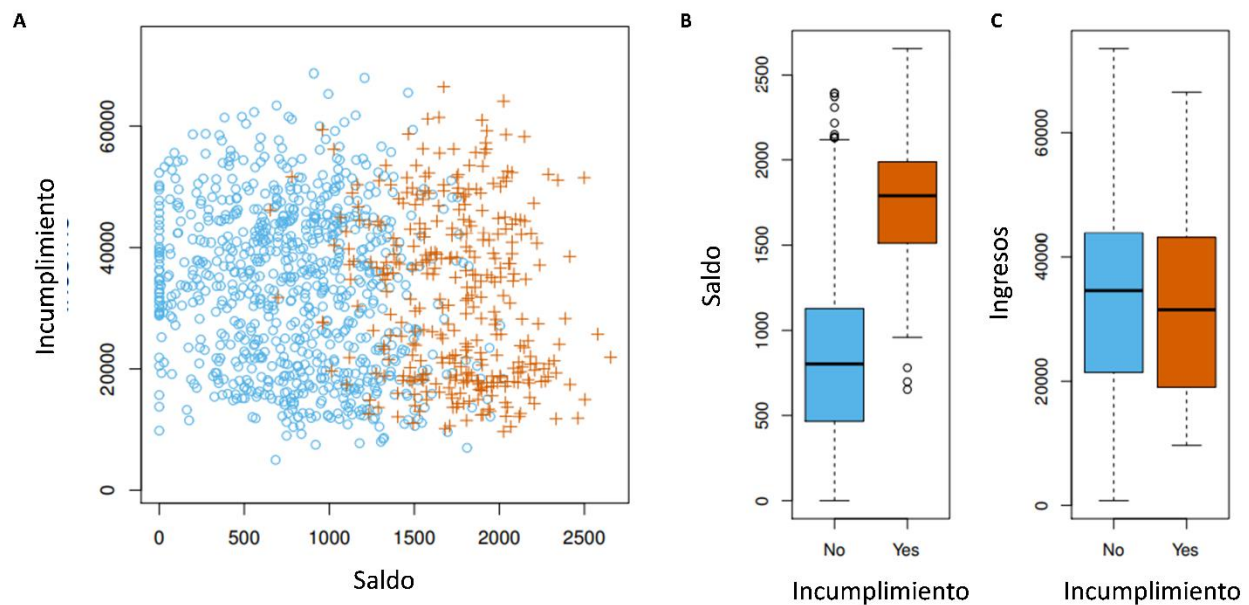


Figura A.1. El conjunto de datos de *Incumplimiento*. **A** – Ingresos anuales en función del saldo mensual de las tarjetas de crédito de 10000 individuos. Los individuos que incumplieron el pago de sus tarjetas de crédito se marcaron en naranja, y aquellos que no lo hicieron se muestran en azul. Obtenido de (James et al., 2013). **B** – Boxplots de *Saldo* como función del estado de *Incumplimiento*. **C** – Boxplots de *Ingresos* como función del estado de *Incumplimiento*.

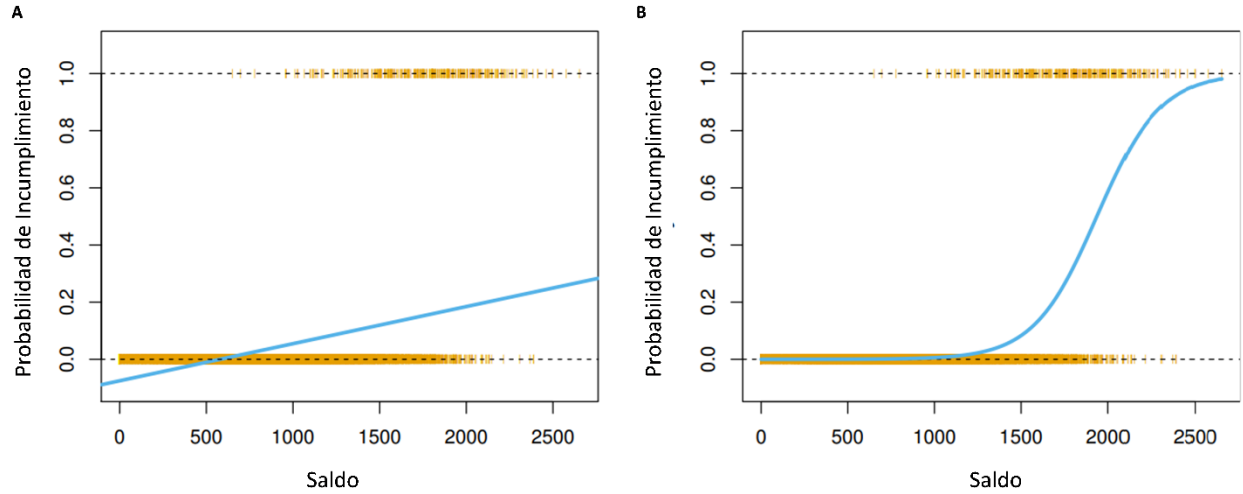


Figura A.2. Probabilidad de *Incumplimiento* en función del *Saldo* usando el conjunto de datos correspondiente obtenido de (James et al., 2013). Las líneas de color naranja representan las observaciones etiquetadas como 0/1. Todas las probabilidades están entre 0 y 1. **A** – Probabilidad de Default estimada mediante una regresión lineal (línea sólida azul). Algunas probabilidades son negativas. **B** – Probabilidad de Default estimada mediante una regresión logística (línea sólida azul). Obtenido de (James et al., 2013).

Por lo tanto, se necesita una función adecuada para que su rango se encuentre dentro del intervalo $[0, 1]$. Esta función puede escribirse como:

$$p(x) = \frac{e^{\beta_0 + B_1 X}}{1 + e^{\beta_0 + B_1 X}} \quad (3)$$

la cual es creciente y se conoce como función logística (**Figura A.2B**). La distribución de la variable respuesta será binomial con el número de observaciones (N) y $p(x)$ como parámetros:

$$Y|X \sim \beta_i(N, p(x)) \quad (4)$$

Los coeficientes se pueden obtener mediante el método de máxima verosimilitud o con el método de descenso por gradiente. Luego, podemos definir un umbral μ tal que $p(x > \mu)$. De esta manera se obtienen las predicciones del modelo y se puede calcular el error de clasificación para determinar su efectividad.

Para incrementar la robustez y efectividad del modelo, se le aplican dos métodos de regularización conocidos como Lasso y Ridge, en un arreglo denominado “Red Elástica”. Esto reduce la varianza, minimiza el riesgo de memorización, atenúa el efecto de correlación entre predictores, y reduce la influencia de aquellos predictores que son menos relevantes en el modelo.

A.2 Árboles de Decisión

Los árboles de decisión son algoritmos que resuelven problemas de regresión y clasificación dividiendo el espacio de predictores en una serie de regiones simples, mediante un conjunto de reglas. Tienen una interpretación sencilla y son altamente robustos en comparación con los modelos lineales debido a que operan con pocas restricciones; no se necesitan los supuestos de linealidad, normalidad de los residuos, homogeneidad de la varianza, entre otros. También son menos sensibles a la baja calidad o ausencia de datos, seleccionan los predictores más importantes al generarse durante el entrenamiento, y no necesariamente los utilizan todos en el modelo final. Sirven cuando la variables respuesta y predictores tanto numéricas como continuas y no necesitan variables dummies, aunque en ocasiones incrementan el desempeño del modelo. Sin embargo, también son susceptibles al sobreajuste, a los datos atípicos, y a los datos desbalanceados. Tampoco suelen ser competitivos con las mejores técnicas de aprendizaje supervisado en términos de predicción, y a menudo la combinación de un gran número de árboles puede mejorar notablemente la predicción en detrimento de la interpretabilidad. Para ilustrar los fundamentos del funcionamiento de los árboles de decisión, utilizaremos el conjunto de datos Hitters del paquete ISLR2 de R, para ilustrar el logaritmo del salario de jugadores de Baseball profesionales de las Grandes Ligas, en función de los años de experiencia y los hits convertidos en el año anterior. La **Figura A.3** muestra el gráfico pertinente con el logaritmo del salario en una escala cromática donde los colores más oscuros representan los valores más bajos, mientras los más claros los más altos. La estructura de un árbol de decisión está conformada por varios tipos de nodos conectados por ramas. En la parte superior aparece un nodo raíz o principal, más abajo uno o más nodos internos y al final uno o varios nodos terminales, también conocidos como hojas (**Figura A.4**). Su lectura se realiza de arriba hacia abajo de modo que las hojas se

encuentran en la parte inferior del árbol. En cada uno de los nodos se establece un valor umbral S al predictor que representa, de manera que las observaciones menores a ese valor se establecen a la izquierda del nodo, mientras que las que tienen valores más grandes se sitúan a la derecha. En consecuencia los nodos con umbrales representan predictores y poseen ramas que representan una regla de decisión. Los nodos terminales no poseen umbrales y por tanto tampoco tienen ramas. Los predictores aparecen en los nodos ordenados de arriba abajo según la importancia que tienen al generar las predicciones del modelo.

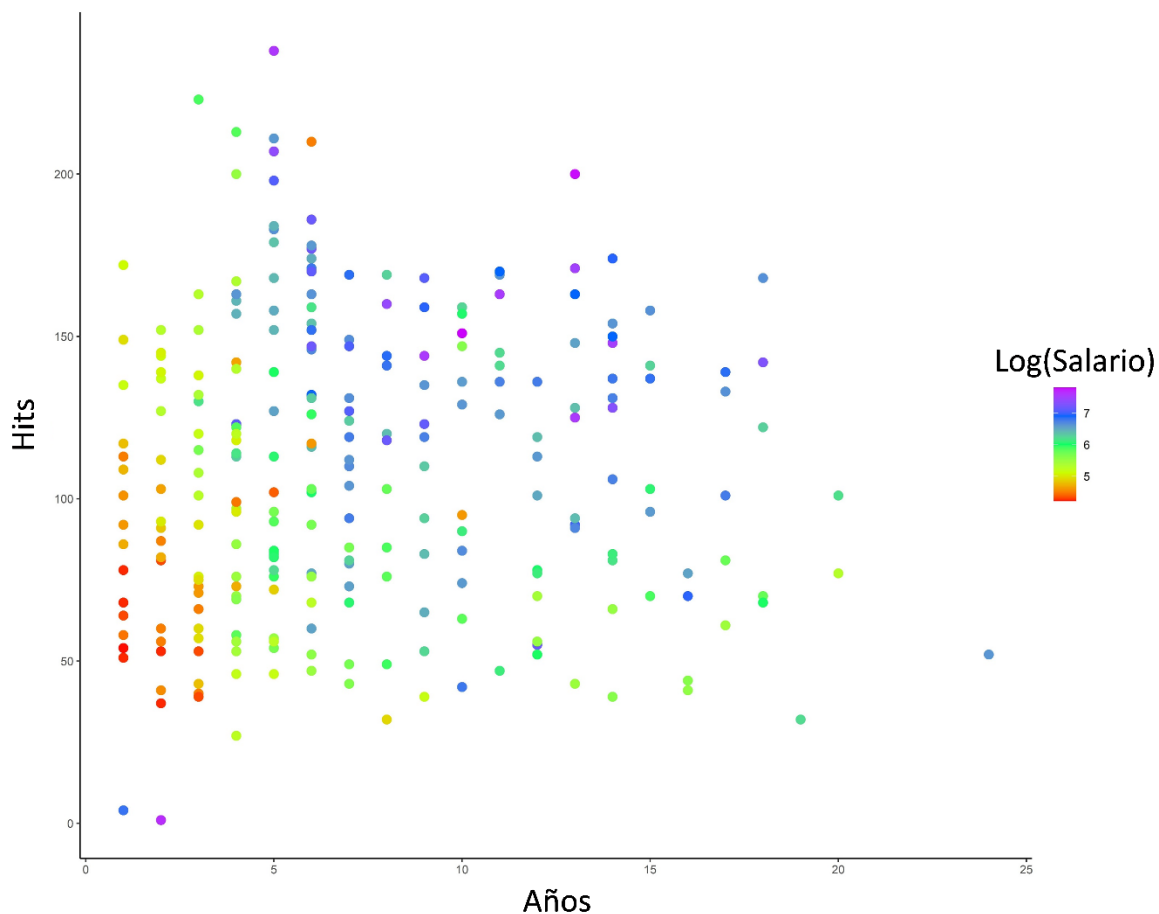


Figura A.3. Datos en los que se construye un árbol de decisión para predecir el logaritmo del salario de jugadores de baseball en las grandes ligas. Se muestra el número de hits que realizaron el año anterior respecto a los años de experiencia. El gradiente de colores corresponde al logaritmo del salario y permite definir intuitivamente umbrales (reglas) de decisión en cada uno de los ejes para formar el árbol.

En la **Figura A.4** se observa que el árbol cuenta con un nodo raíz que representa al predictor Año, seguido por un nodo interno del predictor Hits a la derecha del árbol, y un total de tres hojas o nodos terminales. La estructura final del árbol se determina durante el proceso de entrenamiento, es decir, la configuración del nodo principal, los nodos internos y los umbrales, que a su vez dan lugar a la forma como se segmenta el espacio de las covariables o predictores (**Figura A.5**). Los números de las hojas del árbol representan el valor promedio de las observaciones presentes en cada una de las regiones, es decir, el valor promedio del logaritmo de los salarios.

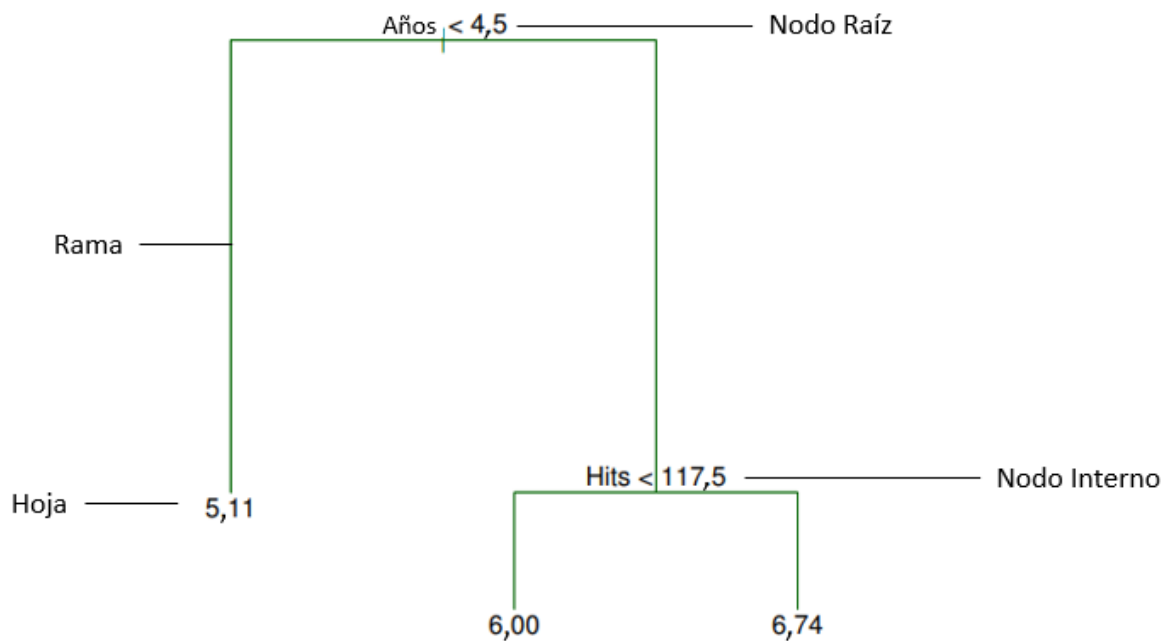


Figura A.4. Árbol de decisión para las variables Años y Hits. Obtenido de (James et al., 2013).

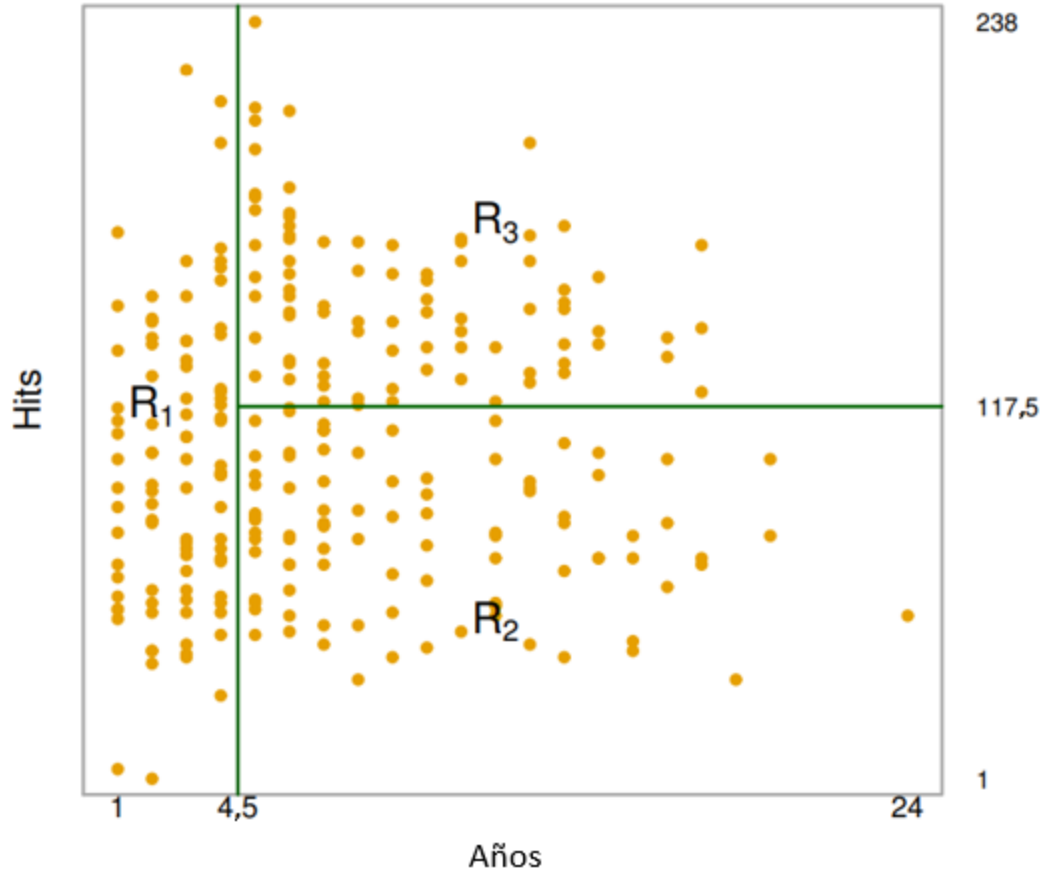


Figura A.5. Segmentación del espacio de predictores en las regiones R1, R2, y R3. Obtenido de (James et al., 2013).

El valor del umbral S en un predictor se determina al dividir el espacio de covariables en dos mitades rectangulares. El objetivo es encontrar las regiones , que minimicen el RSS dado por:

$$RSS = \sum_{j=1}^J \sum_{i \in R_j} (y_i - \widehat{y}_{R_j})^2 \quad (5)$$

donde $\widehat{y}_{R_j}$ es la respuesta media de las observaciones de entrenamiento dentro de la j -ésima caja o región.

En el caso particular de los jugadores, se observa que el umbral de la covariable “Año” es 4,5 y el de la covariable “Hit” es 117,5. Esto tiene sentido porque nos dice que los jugadores tienen un salario relativamente bajo independientemente de los Hits que hayan convertido el año anterior, cuando tienen menos de 4,5 años de experiencia. Al sobrepasar esta cantidad de años, los salarios aumentan notablemente y la cantidad de Hits se hace más relevante. El

algoritmo es capaz de conseguir estos puntos de corte porque, tras probar con varios valores para el umbral, el valor 4,5 para el predictor Año implicó la mayor reducción posible del *RSS*, y lo mismo ocurrió con el valor 117,5 para el predictor Hits. Es decir, para escoger la covariable que representará el nodo principal determina la suma del *RSS* para las dos regiones del espacio que se forman con el umbral, para una covariable particular, y luego establece el umbral que produjo el menor valor. Hace lo mismo para el resto de las covariables y se queda con aquella que produjo el menor valor. Luego desciende un nivel hasta el siguiente nodo y repite el procedimiento, estableciendo las covariables más importantes para cada nivel. Eventualmente el árbol sobreajusta a menos que se limite el número de nodos o el número de observaciones que pueden estar presentes en una hoja. Los caminos del árbol se forman a partir del nodo principal hasta una hoja particular, siendo el valor de esta la predicción del modelo para una observación dada. Por tanto, el estimador de la variable respuesta o dependiente se puede escribir de la forma:

$$\hat{f}(x) = \sum_{m=1}^M \hat{c}_m I(x \in R_m) \quad (6)$$

Donde I es la función indicadora que vale cero en todas las regiones distintas de m .

La descripción anterior se corresponde con un árbol de decisión de regresión, pues la variable respuesta es continua. Sin embargo, veremos más adelante que el problema que intentaremos resolver en este capítulo es de clasificación, donde la variable respuesta es binomial. Por lo tanto, se deberá utilizar una medida distinta al *RSS*, como por ejemplo, el índice Gini o la entropía. No obstante el resto del procedimiento es completamente análogo.

A.3 Ensamblés

Los ensambles son arreglos de estimadores o modelos cuya respuesta es una combinación particular de las respuestas de los modelos que contiene. Los estimadores pueden ser modelos simples o complejos, tener arquitecturas diferentes, recibir conjuntos de datos diferentes, o ser completamente iguales. Cuando se integran en el ensamble producen a menudo mejores respuestas predictivas. Sin embargo, entrenar un ensamble requiere más recursos computacionales y su interpretabilidad se reduce notablemente. Entre los ensambles más

conocidos tenemos a los de Votación, Agregación Bootstrap o Bagging, Bosques Aleatorios y Gradient Boosting.

A.3.1 Votación

El ensamble de votación se utiliza en los problemas de clasificación y consiste en ofrecer una predicción basada en la respuesta que más se repite en los estimadores base (**Figura A.6**). Este método de ensamblado se utiliza sólo cuando los estimadores base o los modelos son diferentes.

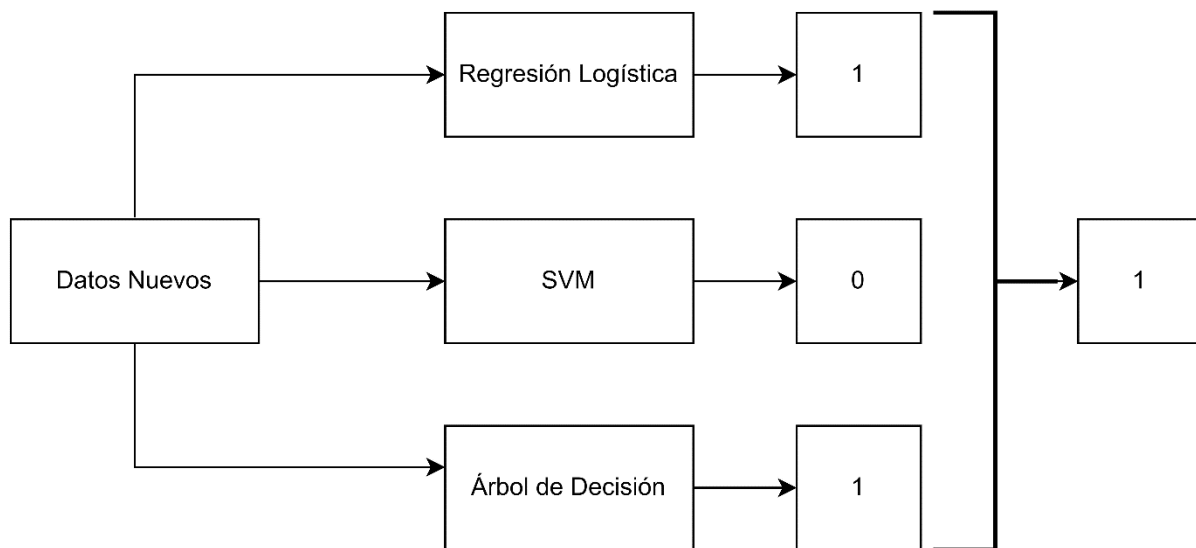


Figura A.6. Esquema de un ensamble de modelos de tipo Votación por Mayoría.

Si los modelos tienen como salida una probabilidad, es posible asignarles un peso a sus respuestas de acuerdo a la cercanía que tengan con la probabilidad cero o uno. Este procedimiento se conoce como *Soft Voting*.

A.3.2 Bagging

El ensamble de Agregación Bootstrap (Baggin) se puede usar tanto con modelos iguales como con modelos distintos. Esto se debe a que para la muestra original de datos se genera un

nuevo conjunto de muestras a partir de una extracción con reemplazo de las observaciones (**Figura A.7**). La calidad de la muestra bootstrap dependerá del número de observaciones, y cada modelo puede recibir una muestra particular. Debido a la naturaleza matemática de las muestras bootstrap, $2/3$ de las observaciones estarán repetidas en cada muestra, por lo que las observaciones restantes pueden usarse como datos de validación durante el entrenamiento de los modelos. Las salidas de cada uno de ellos se promedia y se obtiene la respuesta predictiva del ensemble.

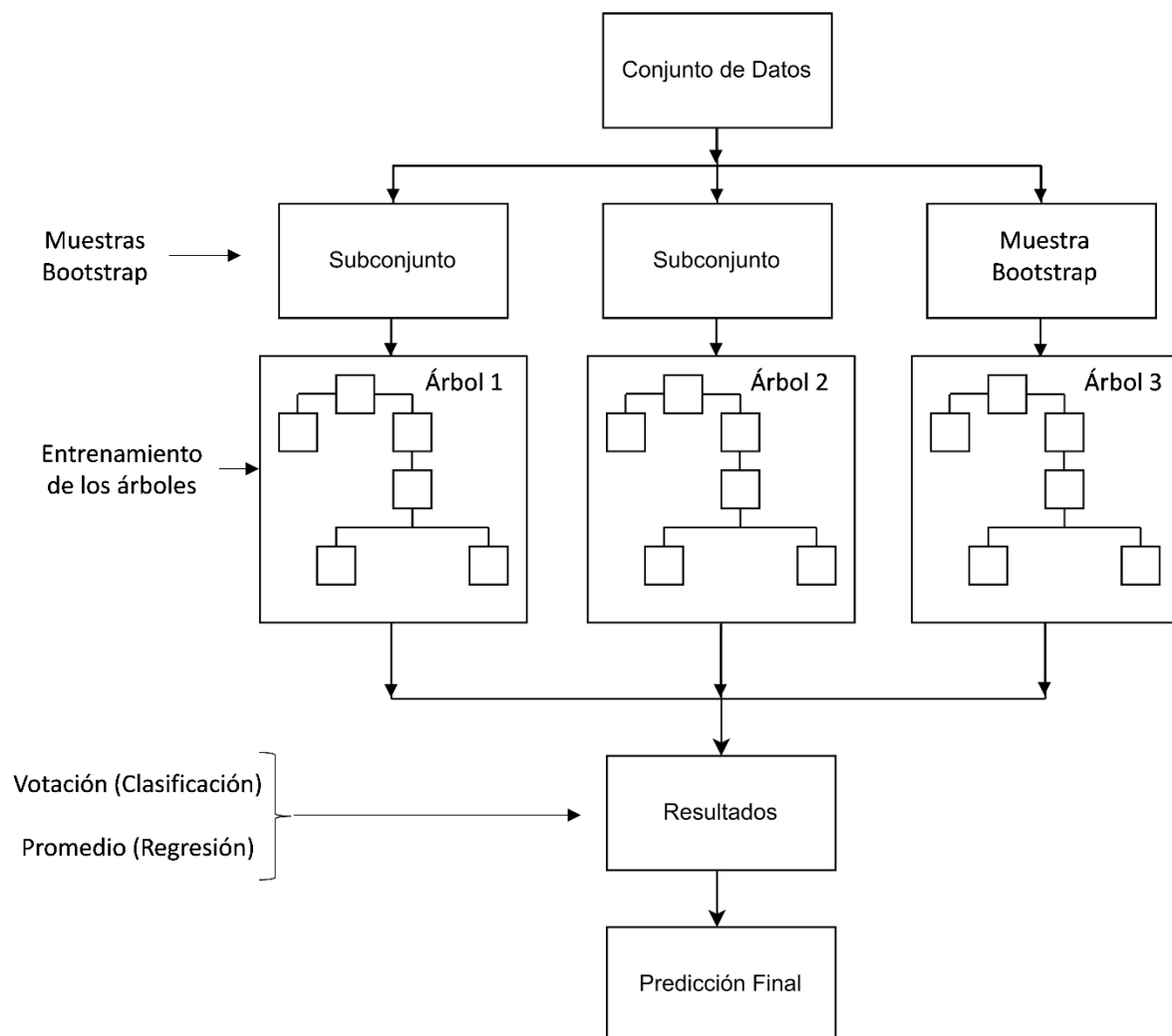


Figura A.7. Esquema de un ensemble de modelos de tipo Bagging.

A.3.3 Random Forest

Este tipo de ensamble se utiliza normalmente con árboles de decisión como estimadores base. El algoritmo es muy similar al de Bagging pero con la diferencia de que se anulan algunos predictores en forma aleatoria en las muestras bootstrap (**Figura A.8**). Esto decorrelaciona los árboles del ensamble y reduce la varianza que se genera al promediar sus respuestas. La correlación de los árboles se produce porque al ser generados de forma codiciosa, es decir, colocando los predictores más importantes al comienzo de cada árbol, existe la posibilidad de que estos predictores de alta importancia aparezcan en los mismos niveles de varios árboles, rompiéndose así la independencia entre ellos. La cantidad de predictores que se toma en cuenta para ser elegido en cada nodo es entonces menor o igual al número total de predictores p , es decir, $m \leq p$, y normalmente, se escoge $m = \sqrt{p}$.

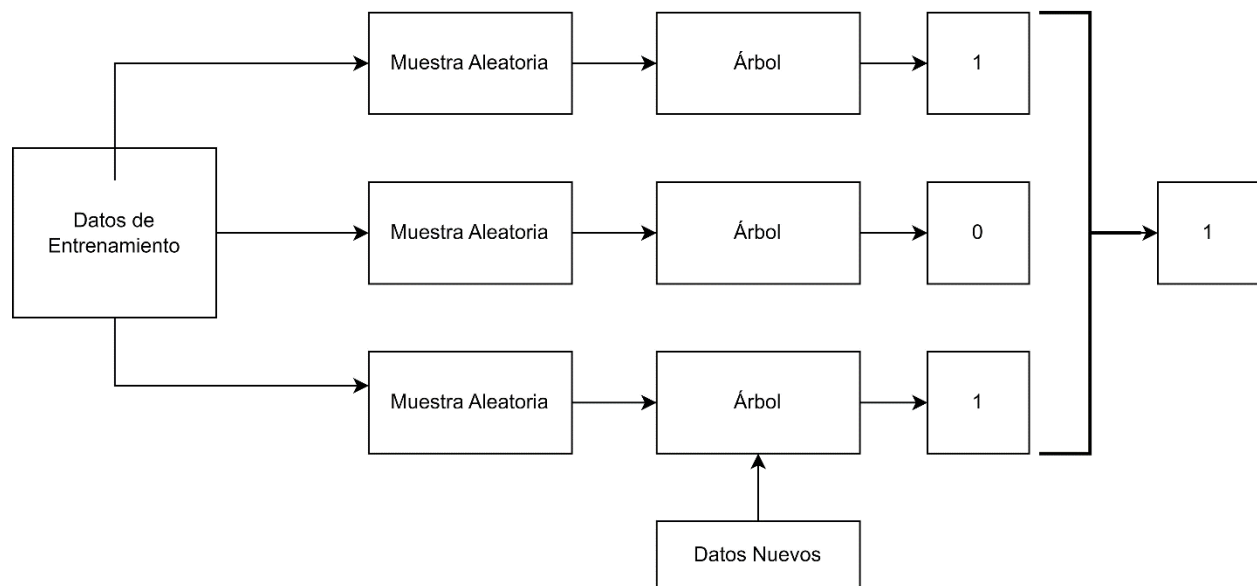


Figura A.8. Esquema de un ensamble de modelos de tipo Bosques Aleatorios (Random Forest).

A.3.4 Gradient Boosting

Este tipo de ensamble implica sumar un conjunto de modelos (*Weak Learners* con poca capacidad predictiva) que aprenden de sus residuos para crear un modelo fuerte capaz de

generar una buena predicción (**Figura A.9**). Los modelos por lo general son árboles de decisión pero también pueden ser de otro tipo. A diferencia de los ensambles de Bagging o Bosques Aleatorios, el Boosting aplica los árboles de forma secuencial mientras la respuesta de cada uno se pesa con un hiperparámetro λ . El algoritmo comienza con un modelo nulo cuya predicción produce un residuo de magnitud igual a la observación real. Es decir, fijamos $\hat{f}(x) = 0$, y $r_i = y_i$ para todos los i del conjunto de entrenamiento, donde r_i es el residuo generado por la predicción del modelo $\hat{f}(x)$. Sea $b = 1, 2, 3, \dots, B$ el índice de los modelos que se van a generar. Se ajusta un árbol con d cortes al conjunto de entrenamiento (X, r) , donde d es el número de nodos no terminales (El número de nodos terminales u hojas siempre será $d + 1$). Luego se actualiza $\hat{f}$ al agregar la predicción de este árbol pesada por el valor de lambda:

$$\hat{f}(x) = \hat{f} + \lambda \hat{f}^b \quad (7)$$

El árbol debería generar residuos menores que el modelo nulo. Luego se actualiza el residuo con la siguiente expresión:

$$r_i = r_i - \lambda \hat{f}^b \quad (8)$$

Después de las b iteraciones, devolvemos el modelo siguiente:

$$\hat{f}(x) = \sum_{b=1}^B \lambda \hat{f}^b \quad (9)$$

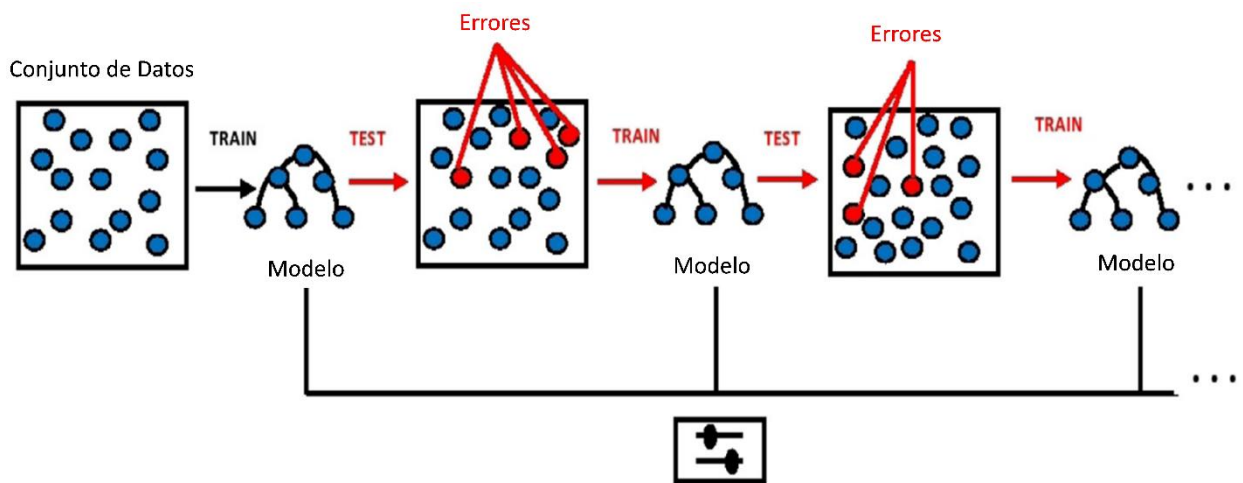


Figura A.9. Esquema de funcionamiento del algoritmo del ensamble de tipo Boosting. Extraído de (Ramzai, 2019).

A.4 Máquinas de Soporte Vectorial (SVM)

Las máquinas de soporte vectorial son modelos que fueron creados para resolver problemas de clasificación, aunque más tarde se descubrió su potencial para resolver problemas de regresión. Para entender los fundamentos de este tipo de modelos, resolveremos un problema de clasificación simple en el que disponemos de observaciones pertenecientes a dos clases linealmente separables (**Figura A.10**). En este espacio bidimensional (con dos predictores), las clases pueden separarse mediante una recta que pase entre el grupo de observaciones de cada una (**Figura A.11A**). Nótese que si se tienen p predictores el espacio tendrá p dimensiones, mientras que la recta será un hiperplano separador de $p - 1$ dimensiones.

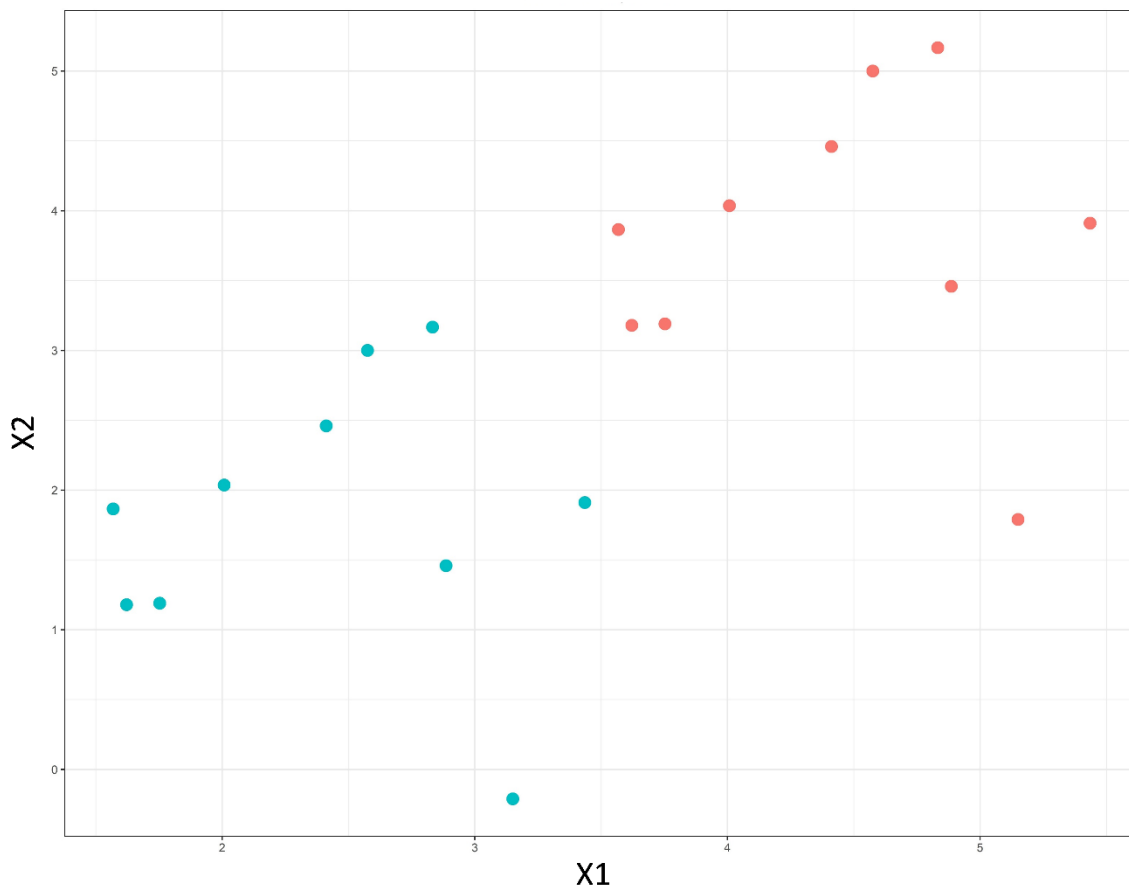


Figura A.10. Dos clases de observaciones linealmente separables.

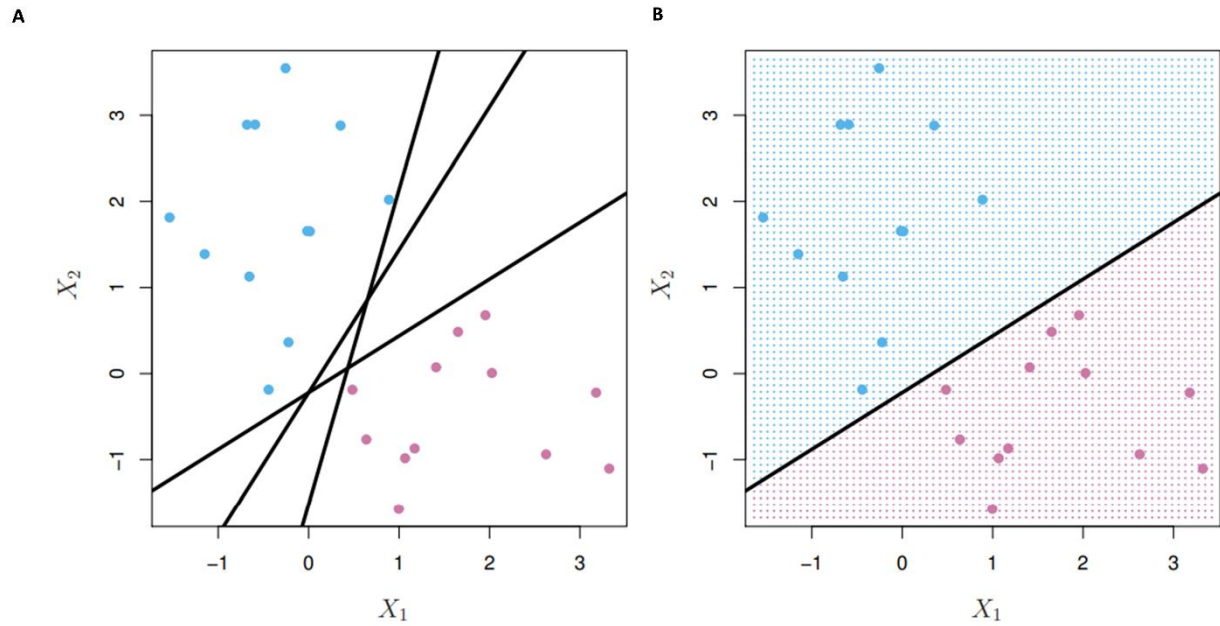


Figura A.11. Hiperplanos de separación para un problema de clasificación de dos clases. **A** - Posibles hiperplanos de separación entre las dos clases. **B** – Hiperplano de separación de Margen Maximal. La zona azul y la zona roja indican la regla de decisión obtenida mediante un clasificador basado en la en este hiperplano de separación. Una observación que caiga en la zona azul será asignada como perteneciente a la clase azul, y de forma equivalente, una observación que caiga en la zona roja se asignará a la clase roja. Extraído de (James *et al.*, 2013).

Existen infinitos hiperplanos que dividen a los grupos de observaciones, por lo que se escoge aquel que separa mejor las clases maximizando la distancia de las observaciones más cercanas al hiperplano (**Figura A.11B**). Matemáticamente tenemos n observaciones X con p predictores, e.g., $X: x_j \in \{x_1, x_1, \dots, x_p\}$, en un espacio $\mathbb{R}^p$, y una respuesta que viene dada por $y_i \in \{1, -1\}$ con $C_1 = \{y_i = 1\}$, $C_2 = \{y_i = 2\}$, siendo C_i la clase i . La ecuación que define al hiperplano en el espacio $\mathbb{R}^p$ puede escribirse como:

$$\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = 0 \quad (10)$$

Asumiendo que para nuestro caso bidimensional este hiperplano divide el espacio $\mathbb{R}^p$ en dos regiones, tenemos que la observación X está en el hiperplano si se cumple la igualdad en (1), y pertenece a una determinada clase según lo siguiente:

$$X_i \in C_1 \Leftrightarrow y_i = 1 \Leftrightarrow \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p > 0 \quad (11)$$

$$\Leftrightarrow y_i(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) > 0 \quad (12)$$

$$X_i \in C_2 \Leftrightarrow y_i = -1 \Leftrightarrow \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p > 0 \quad (13)$$

$$\Leftrightarrow y_i(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) > 0 \quad (14)$$

Esto significa que y_i y $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ tienen el mismo signo, por lo que (2) y (3) pueden expresarse mediante la siguiente expresión:

$$y_i(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) > 0 \quad (15)$$

Por lo tanto, si el hiperplano existe, clasificamos evaluando el signo de la expresión $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ para denotar la clase a la que pertenece la observación. Para encontrar el hiperplano de separación, que denotamos π de ahora en adelante, es necesario determinar los valores de los coeficientes β_i . Nótese que el vector formado por estos coeficientes es ortogonal a π :

$$\boldsymbol{\beta} = (\beta_0, \dots, \beta_p) \perp \pi \quad (16)$$

La ecuación de la recta que pasa por el punto P y es perpendicular al hiperplano π , viene dada por:

$$\mathbf{r}(t) = t\boldsymbol{\beta} + \mathbf{P} \quad (17)$$

El parámetro t puede determinarse asumiendo que $P = (x_1, x_2, \dots, x_p)$ pertenece al plano π , e.g., $t\boldsymbol{\beta} + \mathbf{P} \in \pi$:

$$\beta_0 + \sum_{i=1}^p \beta_i(t\beta_i + x_i) \quad (18)$$

$$\beta_0 + \sum_{i=1}^p t\beta_i^2 + \sum_{i=1}^p \beta_i x_i = 0 \quad (19)$$

$$\beta_0 + t\|\boldsymbol{\beta}\|^2 + \boldsymbol{\beta} \cdot \mathbf{P} = 0 \quad (20)$$

$$t = -\frac{\beta_0 + \boldsymbol{\beta} \cdot \mathbf{P}}{\|\boldsymbol{\beta}\|^2} \quad (21)$$

Entonces, la distancia de cualquier punto al plano viene dada por:

$$d(t\boldsymbol{\beta} + \mathbf{P}) = \|t\boldsymbol{\beta} + \mathbf{P} - \mathbf{P}\| = |t|\|\mathbf{P}\| = \frac{|\beta_0 + \boldsymbol{\beta} \cdot \mathbf{P}|}{\|\boldsymbol{\beta}\|} = \frac{1}{\|\boldsymbol{\beta}\|} |\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p| \quad (22)$$

Si asumimos que $\|\boldsymbol{\beta}\| = 1$ nos queda que:

$$d(t\boldsymbol{\beta} + \mathbf{P}, \mathbf{P}) = |\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p| \quad (23)$$

En consecuencia, el módulo de la ecuación del hiperplano determina la distancia del punto P a π . Luego, el signo de la ecuación del hiperplano nos determina la clase a la que pertenece la observación, pero al mismo tiempo el módulo nos da información acerca de qué tan lejos o cerca se encuentra el punto del hiperplano. La función f que define al hiperplano viene dada por:

$$f(x) = f(x_1, x_2, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = \beta_0 + \sum_{i=1}^p \beta_i x_i \quad (24)$$

Por lo tanto, para encontrar el vector β del hiperplano separador más óptimo se maximiza la distancia M de los puntos (observaciones) más cercanos al hiperplano, los cuales se denominan vectores de soporte (**Figura A.12**). En consecuencia, el problema de optimización queda definido de la siguiente forma:

$\max M_{\beta_0, \beta_1, \dots, \beta_p, M}$, restringido a:

$$\sum_{j=1}^p \beta_j^2 = 1 \quad (25)$$

$$y_i(\beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i) > M \text{ para todo } i \quad (26)$$

El problema tiene solución para $M > 0$ si existe un hiperplano separador. El valor óptimo de M es la medida del margen de las bandas de separación (hiperplanos paralelos al hiperplano separador que contienen a los puntos de soporte). La existencia del hiperplano separador conforma el Clasificador de Margen Maximal, que funciona para el caso particular en el que las clases son completamente separables linealmente (**Figura A.12**). Sin embargo, en casos reales esta condición no suele cumplirse debido a la existencia de observaciones que no pueden separarse mediante el hiperplano (**Figura A.13**). Por lo tanto, es necesario realizar algunas modificaciones al problema de optimización, de forma que se admita un pequeño porcentaje de error al clasificar las observaciones, una vez definido el valor de M para los márgenes (**Figura A.14**).

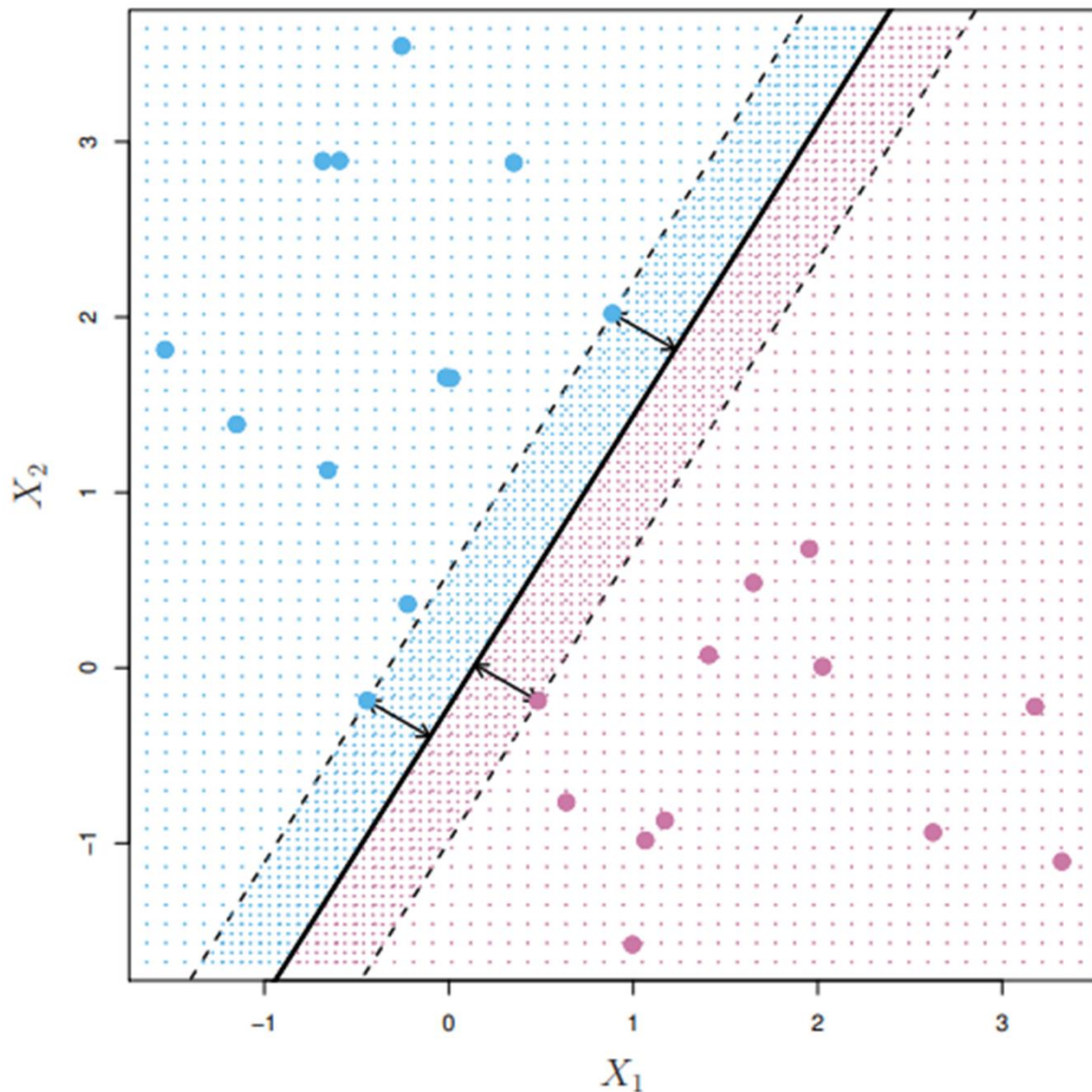


Figura A.12. Hiperplano separador con vectores de soporte. El hiperplano separador de máximo margen se muestra con una línea sólida. El margen es la distancia desde la línea sólida hasta cualquiera de las líneas punteadas. Los dos puntos azules y el punto rojo que se encuentran sobre los márgenes son los vectores de soporte y la distancia de estos puntos al hiperplano se indica con flechas. Las grillas azul y rojas indican la regla de decisión adoptada por el clasificador basada en el hiperplano separador. Extraído de (James *et al.*, 2013).

Cuando se incluyen estas modificaciones el clasificador se denomina Clasificador de Soporte Vectorial, y la ecuación (25) se transforma en:

$$y_i(\beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i) > M(1 - \varepsilon_i) \text{ para todo } i \quad (27)$$

Donde ε_i es una variable que nos permite modificar el nivel de aceptación de los errores de clasificación:

- i) $\varepsilon_i = 0$ significa que las clasificaciones correctas se producen fuera o sobre el margen M .
- ii) $0 < \varepsilon_i < 1$ significa que habrá observaciones clasificadas como correctas dentro de los márgenes.
- iii) $\varepsilon_i > 1$ indica que habrá observaciones clasificadas incorrectamente.

Para controlar el error permitido se utiliza un hiperparámetro C que cumple la siguiente condición:

$$\sum_i \varepsilon_i \leq C \quad (28)$$

Si $C = 0$, entonces $\varepsilon_i = 0$ y el problema de optimización es equivalente a (11). Si $C \leq 1$ entonces $\varepsilon_i < 1$ para todo i , lo que significa que las observaciones clasificadas como correctas por encima de un margen inferior a M . Por lo tanto, asumiendo que $C > 0$, y que $\varepsilon_i \geq 0$, tenemos que el problema de optimización resulta ser:

$\max M_{\beta_0, \beta_1, \dots, \beta_p, \varepsilon_1, \dots, \varepsilon_n, M}$, restringido a:

$$\sum_{j=1}^p \beta_j^2 = 1 \quad (29)$$

$$y_i(\beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i) \geq M(1 - \varepsilon_i) \text{ para todo } i \quad (30)$$

$$\sum_i \varepsilon_i \leq C \quad (31)$$

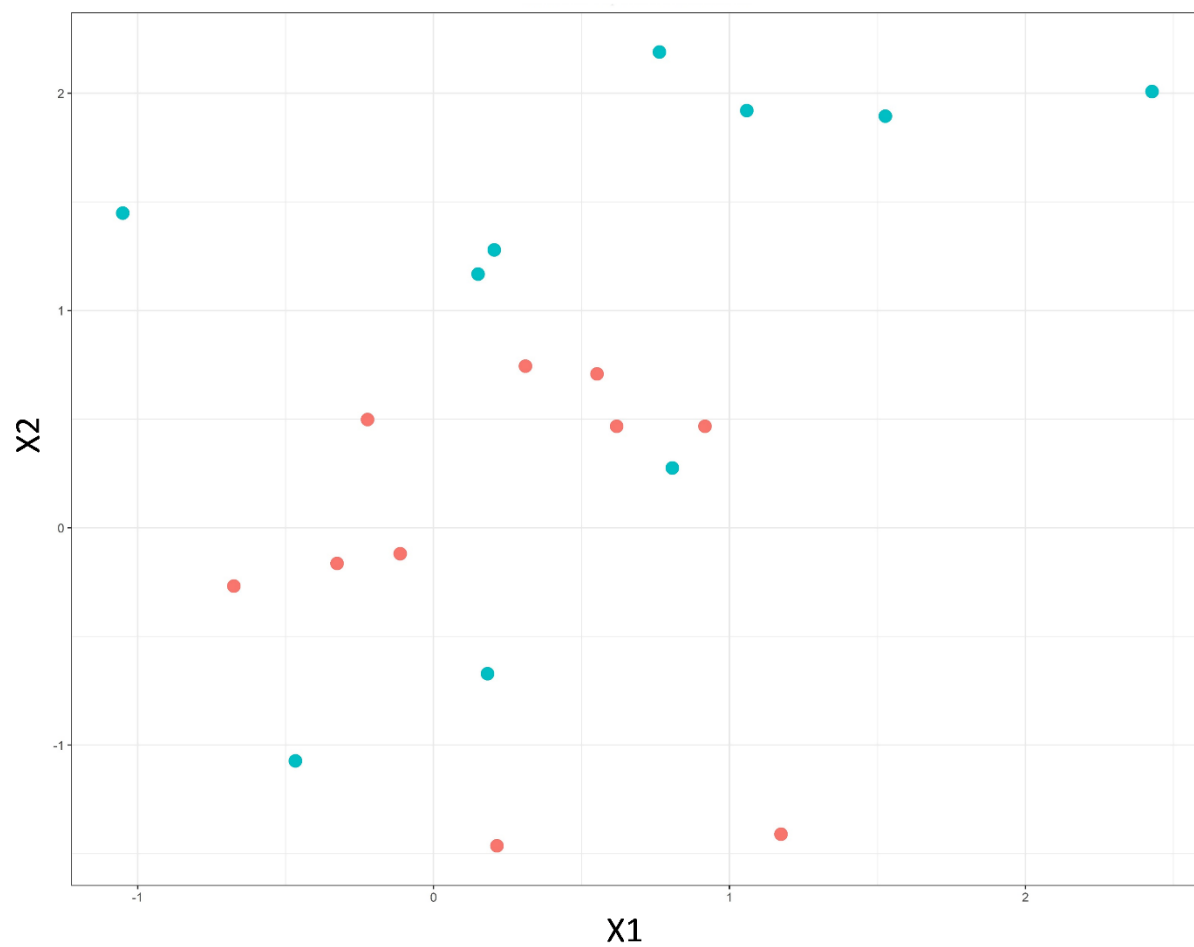


Figura A.13. Dos clases de observaciones linealmente no separables.

Un valor alto de C implica un margen grande que engloba un número elevado de observaciones para clasificar (**Figura A.14A**). Esto deriva en baja varianza y alto sesgo. Por el contrario, si el valor de C disminuye, el sesgo se reduce y se incrementa la varianza (**Figura A.14B-D**).

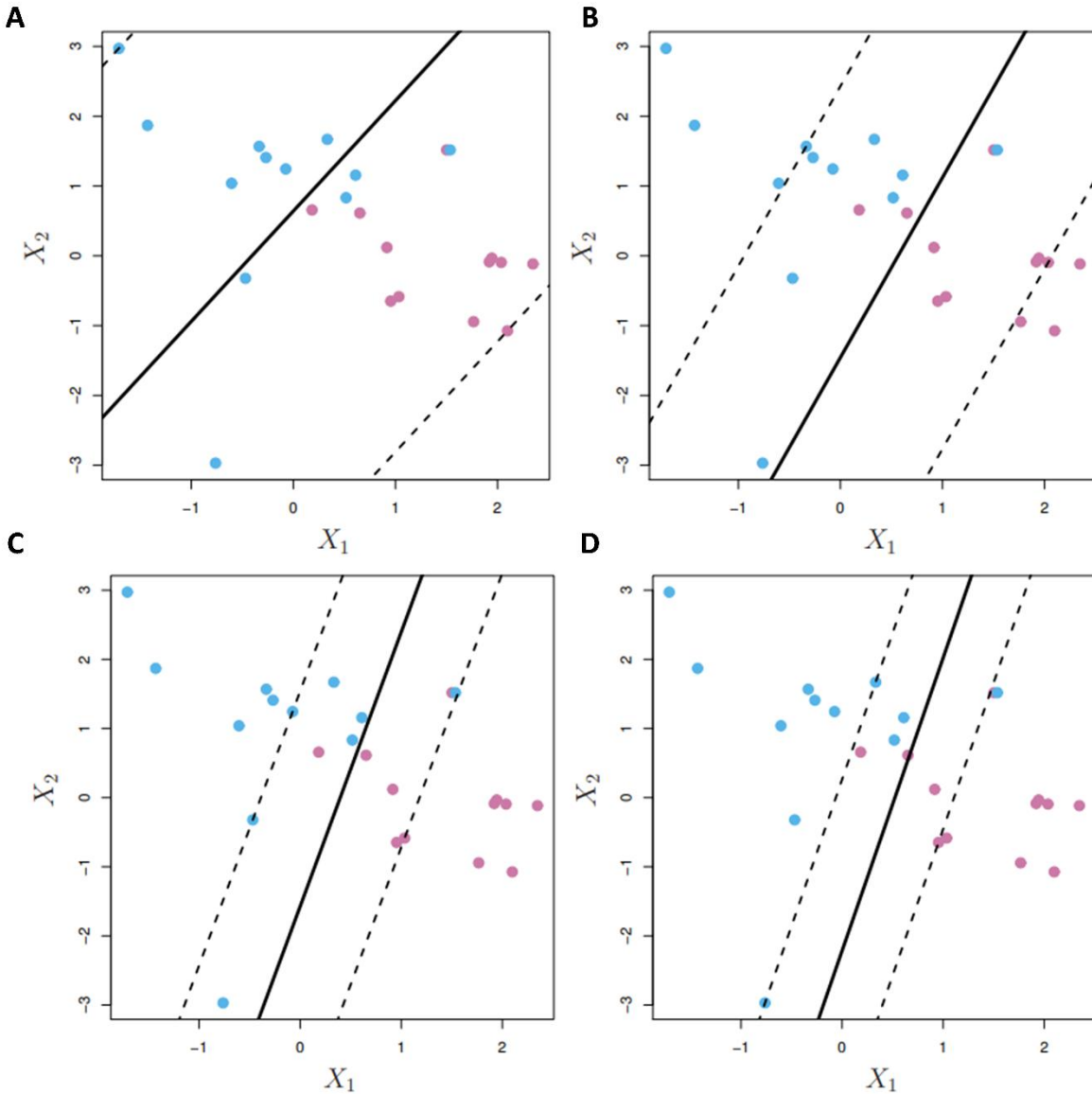


Figura A.14. Un clasificador de soporte vectorial se ajustó con cuatro valores diferentes del hiperparámetro de calibración. El valor más grande de C se usó en **A**, mientras que valores cada vez menores se usaron en **B**, **C** y **D**. Cuando C es grande, existe una gran tolerancia de observaciones que caen dentro del margen, y por tanto dicho margen será grande. A medida que C decrece, la tolerancia de las observaciones que caen en el lado equivocado del margen se reduce, y el margen se hace más estrecho. Extraído de (James *et al.*, 2013).

Finalmente, las Máquinas de Soporte Vectorial (SVMs, por sus siglas en inglés) son los clasificadores más generales, debido a que utilizan kernels (K) que modifican el espacio de

predictores para establecer fronteras de clasificación no lineales que permiten resolver problemas de clases que no son separables linealmente. Esto se logra al escribir la ecuación (9) en términos de las observaciones $x^i = (x_1^i, \dots, x_p^i)$. Para $i = 1, \dots, n$:

$$\begin{aligned} g(x) &= \beta_0 + \sum_{i=1}^n a_i x \cdot x^i \\ &= \beta_0 + \sum_{i=1}^n a_i (\sum_{j=1}^p x_j x_j^i) \\ &= \beta_0 + \sum_{j=1}^p (\sum_{i=1}^n a_i x_j^i) x_j \end{aligned} \quad (32)$$

Para que $f = g$ debe cumplirse que:

$$\beta_j = \sum_{i=1}^n a_i x_j^i \quad (33)$$

Lo que representa un sistema de p ecuaciones con n incógnitas. En estas condiciones tendremos:

$$f(x) = \beta_0 + \sum_{i=1}^n a_i x \cdot x^i \quad (34)$$

Para obtener los valores de a_i es suficiente computar el producto interno entre $x \cdot x^i$. Luego, mediante el método de máxima verosimilitud y cálculos numéricos, se pueden obtener dichos valores. Podemos llamar $K(x, x^i) = x \cdot x^i$ y obtenemos:

$$f(x) = \beta_0 + \sum_{i=1}^n a_i K(x, x^i) \quad (35)$$

Cuando K representa una frontera lineal, la ecuación (35) nos da una medida de la distancia de la observación a la frontera de clasificación (hiperplano separador). Pero es posible expresar otras funciones del producto punto entre las observaciones que pueden ofrecer fronteras no lineales que cuantifiquen similitud entre dos observaciones. Por lo que la ecuación (19) es una forma más general de la ecuación (9). Algunos de los kernels más utilizados son:

-Kernel lineal:

$$K(x^j, x^i) = x^j \cdot x^i = \sum_{i=1}^p x_j^i x_i^j \quad (36)$$

-Kernel Polinomial:

$$K_d(x^j, x^i) = (c + x^j \cdot x^i)^d \quad (37)$$

-Kernel Gaussiano (RBF):

$$K_{\gamma}(x^j, x^i) = \exp\left(-\gamma \sum_{k=1}^p (x_k^j - x_k^i)^2\right) \quad (38)$$

Los kernels en general incrementan la dimensión del espacio de los predictores. En este espacio de mayor dimensión es posible separar las clases con una frontera de clasificación lineal obtenida al aplicar un SVC lineal (**Figura A.15**), y luego volver al espacio original usando las curvas de nivel como frontera de clasificación (**Figura A.16**).

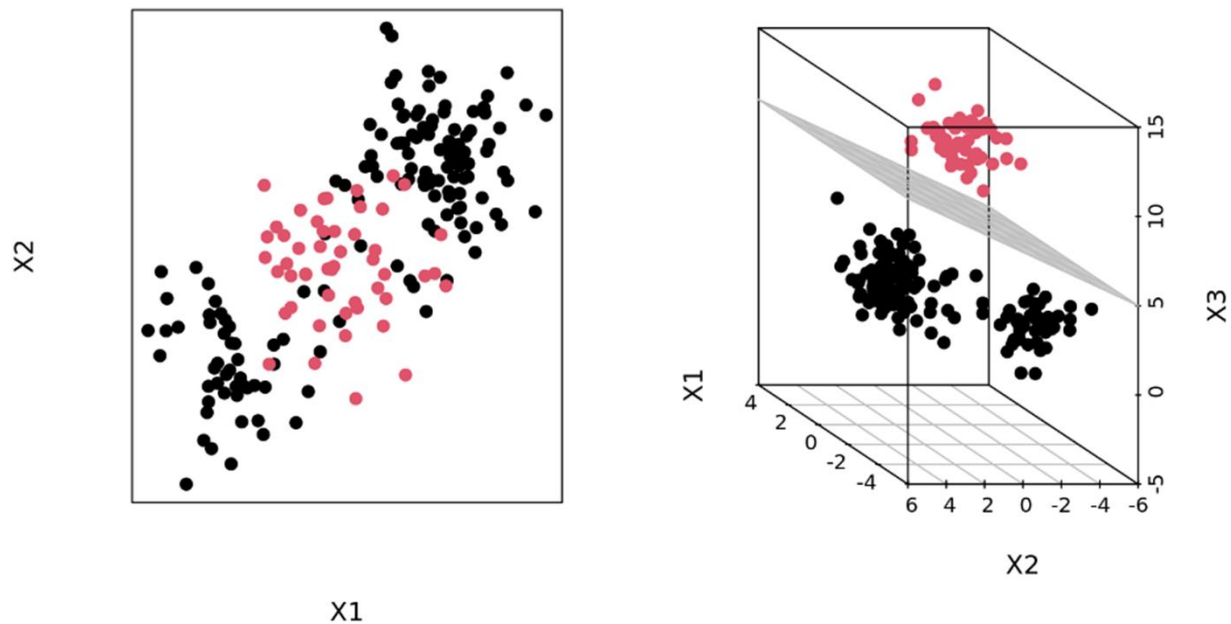


Figura A.15. Dos clases de observaciones linealmente no separables en dos dimensiones se pueden separar en un espacio de predictores de tres dimensiones.

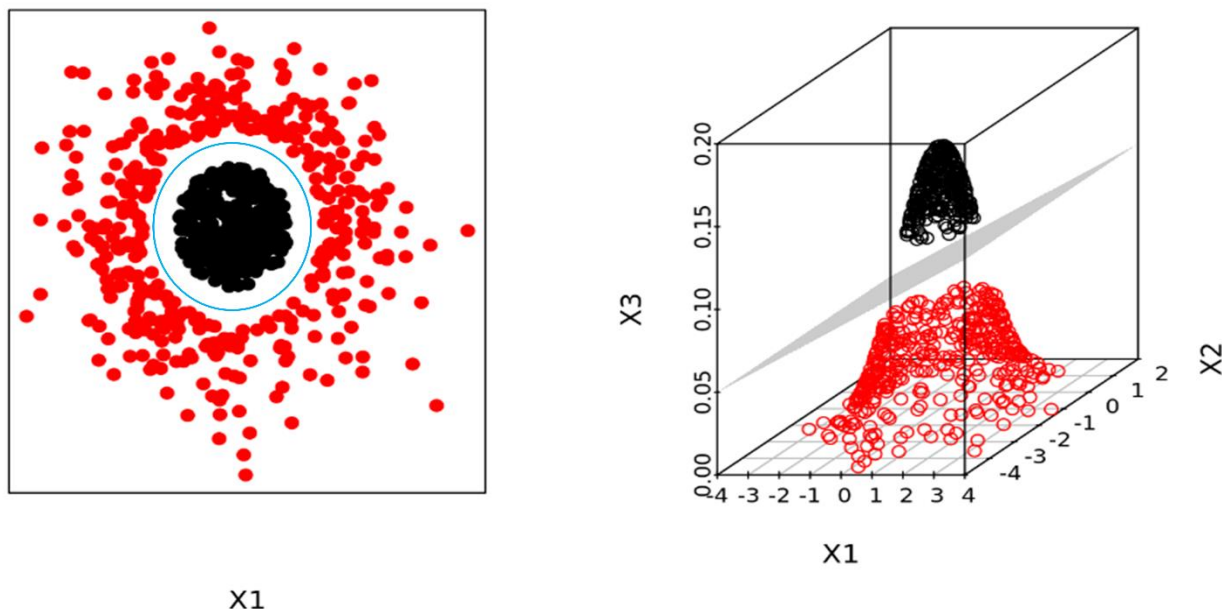


Figura A.16. Dos clases de observaciones linealmente no separables en dos dimensiones se pueden separar en un espacio de predictores de tres dimensiones. El hiperplano separador (lineal), encontrado mediante el entrenamiento de una SVM lineal en tres dimensiones, se transforma en una frontera de clasificación no lineal al proyectarlo en dos dimensiones (Línea sólida azul). Modificado de https://www.cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines.

Al usar el kernel polinomial con $c = 0$ y $d = 1$ se obtiene el kernel lineal (**Figura A.17A**). Con $d > 1$ se generan límites de clasificación no lineales (**Figura A.17B**). En el kernel RBF, se utilizan dos hiperparámetros: γ y C , que es común a todos los kernels (**Figura A.17C**). Valores más altos de γ incrementan la flexibilidad del modelo.

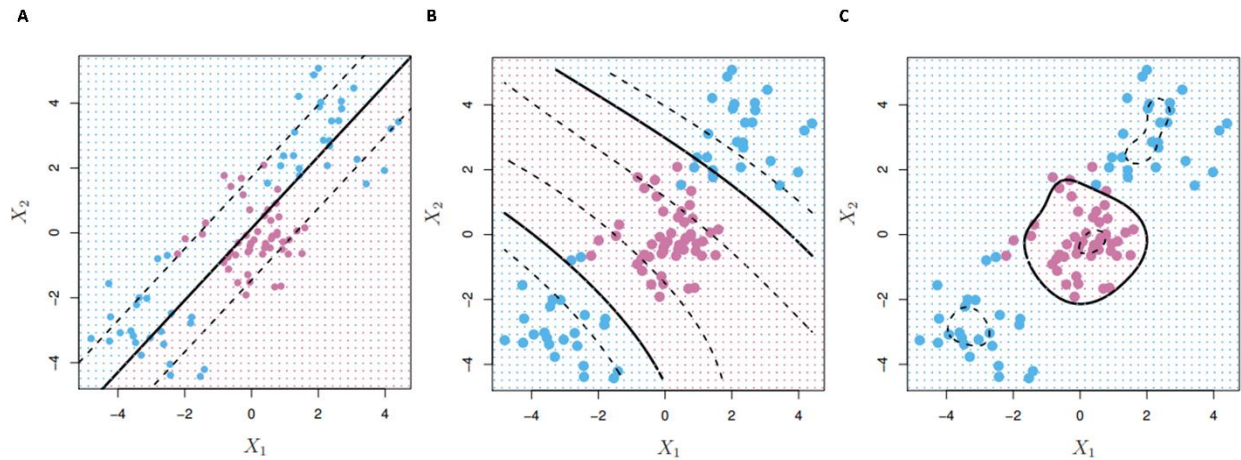


Figura A.17. Fronteras de decisión obtenidas al entrenar un modelo SVM con tres tipos de kernels aplicado en datos no lineales. **A** – Kernel lineal. **B** – Kernel polinomial de grado tres que proporciona una frontera de clasificación más adecuada que el kernel lineal. **C** – Kernel RBF. Extraído de (James *et al.*, 2013).

Existe un gran número de kernels que son más o menos adecuados dependiendo de cada situación particular, cada uno con una serie de hiperparámetros que pueden encontrarse mediante validación cruzada. No obstante, se recomienda probar el kernel RBF, ya que solo posee dos hiperparámetros y su flexibilidad va desde un clasificador lineal a un modelo mucho más complejo.

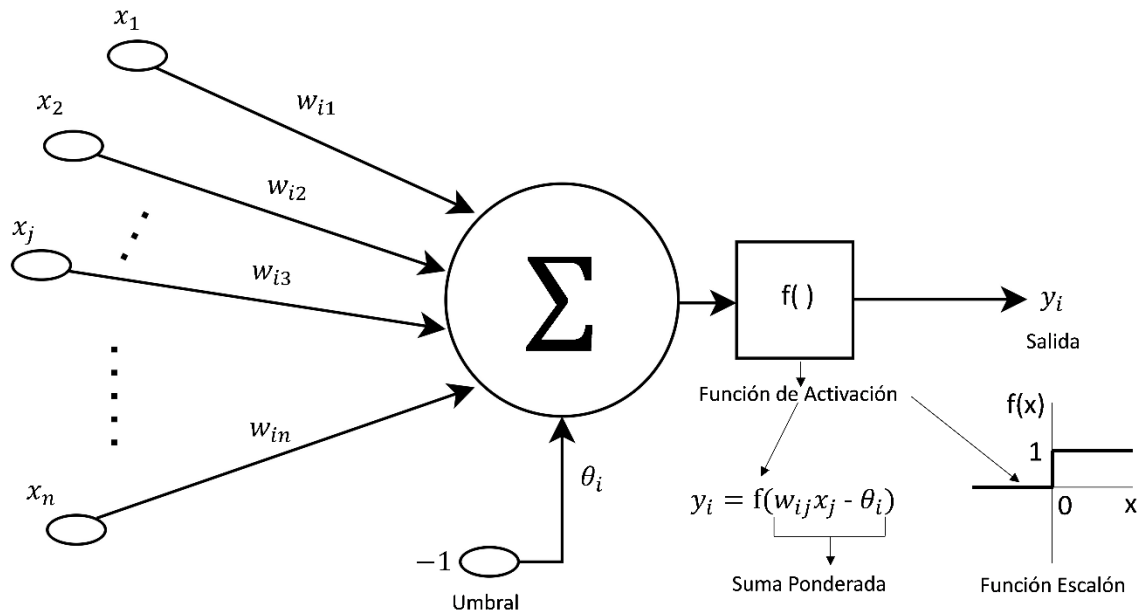
A.5 Redes Neuronales Artificiales

Las redes neuronales artificiales (ANNs, sigla del inglés *Artificial neural networks*) han demostrado ser revolucionarias en las últimas décadas y han jugado un papel de gran relevancia en la resolución de problemas muy diversos. Esto ha sido posible por el advenimiento de ordenadores lo suficientemente potentes para manejar la enorme cantidad de datos que suelen necesitar para su buen funcionamiento.

Las ANNs son modelos matemáticos cuyos elementos fundamentales, las neuronas artificiales (AN), en algunos aspectos tienen un funcionamiento similar a las neuronas biológicas.

Se utilizan principalmente para tratar problemas de extracción y clasificación de características, reconocimiento de patrones y problemas de regresión o *forecasting*. En la **Figura A.18A** se representa una neurona artificial de uno de los tipos más comunes: la de McCulloch y Pitts. Recibe un número determinado de valores llamados entradas, que en nuestro caso pueden representar las muestras que forman un segmento del electromiograma.

A



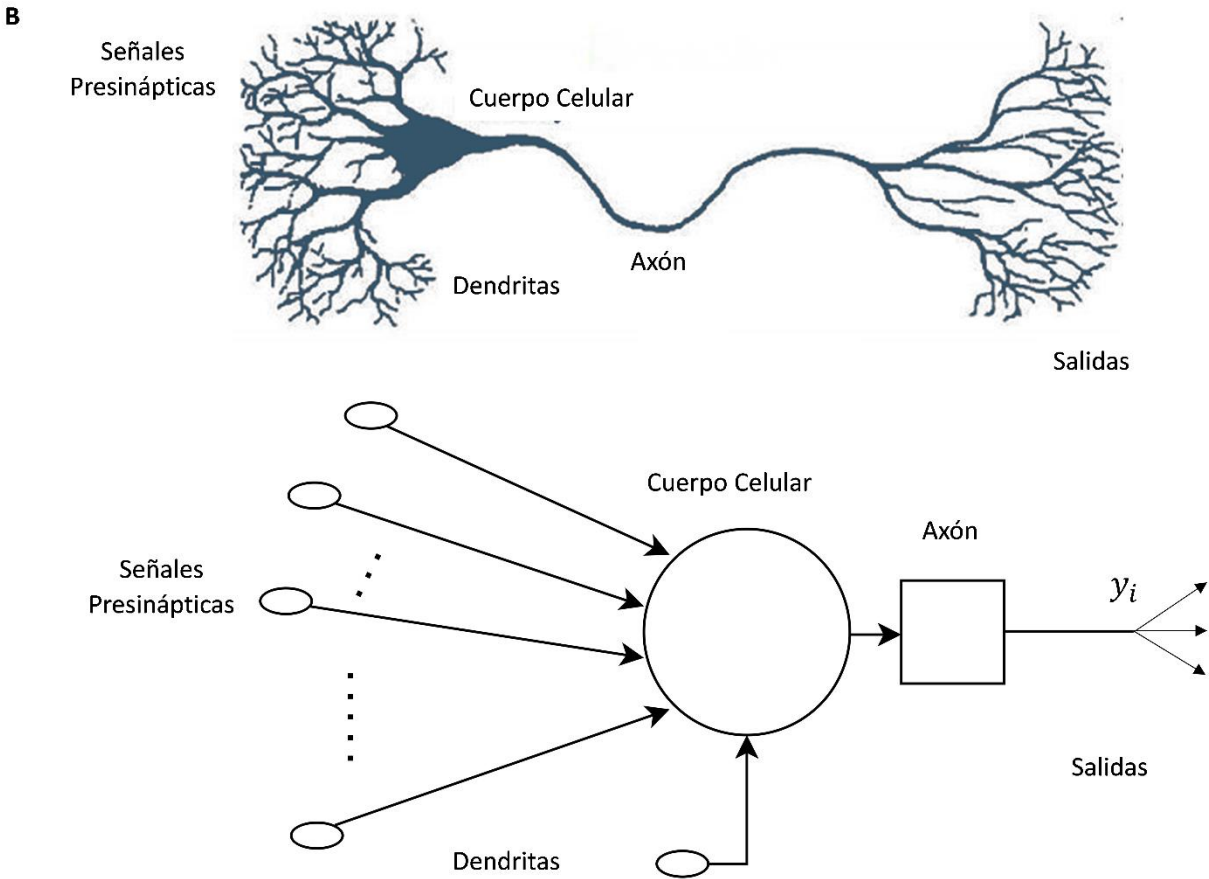


Figura A.18. Diagrama de una neurona de McCulloch-Pitts, unidad fundamental de una red neuronal, y su analogía con una neurona biológica. **A-** Partes de una neurona de McCulloch-Pitts y sus parámetros asociados –entradas (x_i), pesos (w_{ij}), el umbral θ_i , salidas (y_i) y la función de activación de tipo escalón $H(x)$. **B-** Analogía entre una neurona biológica y la neurona artificial de McCulloch-Pitts. Las señales presinápticas pueden representarse por los valores de entrada x_i , pesados por los pesos w_{ij} . La convergencia de dichas señales se representa por la función de activación aplicada a la suma ponderada de las entradas y el valor umbral θ_i , sería el valor mínimo de intensidad que se necesitaría para disparar un potencial de acción en la neurona biológica. La señal que sale de las arborizaciones terminales se representa por las salidas y_i . Imagen de neurona biológica modificada de <https://magiquo.com/redes-neuronales-o-el-arte-de-imitar-el-cerebro-humano/>.

A.5.1 Perceptrón Simple (SP)

Los valores de entrada en una AN están regulados por los pesos w_{ij} . También existe una entrada con valor unitario cuyo peso viene dado por θ_i y se conoce como umbral. La suma ponderada por los pesos de las entradas y el umbral se filtra con una función conocida como

función de activación o función de transferencia. Cuando la función de transferencia o activación es la función escalón, el rango de valores de salida estará entre 0 y 1 dependiendo de si la suma ponderada de las entradas es positiva o negativa.

Finalmente, las salidas y_i se comparan con los valores esperados ya conocidos de antemano y se calcula el error de acuerdo a la diferencia entre ellos. Si las salidas son incorrectas, los pesos θ_i y w_{ij} se modifican en iteraciones sucesivas para minimizar el error hasta que todas las entradas sean clasificadas correctamente o hasta que se alcance un mínimo absoluto de la función de error, que depende de los pesos. Esta tarea se realiza a través de una regla llamada aprendizaje clásico del perceptrón y constituye el proceso de entrenamiento del modelo. Un arreglo de una o varias neuronas artificiales con las características descritas también se le conoce como *Perceptrón Simple* o monocapa.

En la **Figura A.18B** se muestra cómo este modelo se inspira en el funcionamiento de una neurona biológica: las entradas ponderadas son equivalentes a las señales presinápticas con diferentes intensidades que entran por las dendritas y se suman en el cuerpo, mientras que el umbral θ es equivalente al umbral de voltaje necesario para que se produzca un potencial de acción. La función de activación puede representar la presencia o ausencia del potencial de acción (que es el caso de la función escalón), el incremento de la frecuencia, en caso de una función identidad, la imposibilidad de disparar a una frecuencia menor que cero si la función es lineal rectificadora, e incluso el límite en la frecuencia de disparo si la función de activación es de tipo sigmoide. Finalmente, la señal eléctrica que arriba a las arborizaciones terminales que conectan con dendritas de otras neuronas serían las salidas

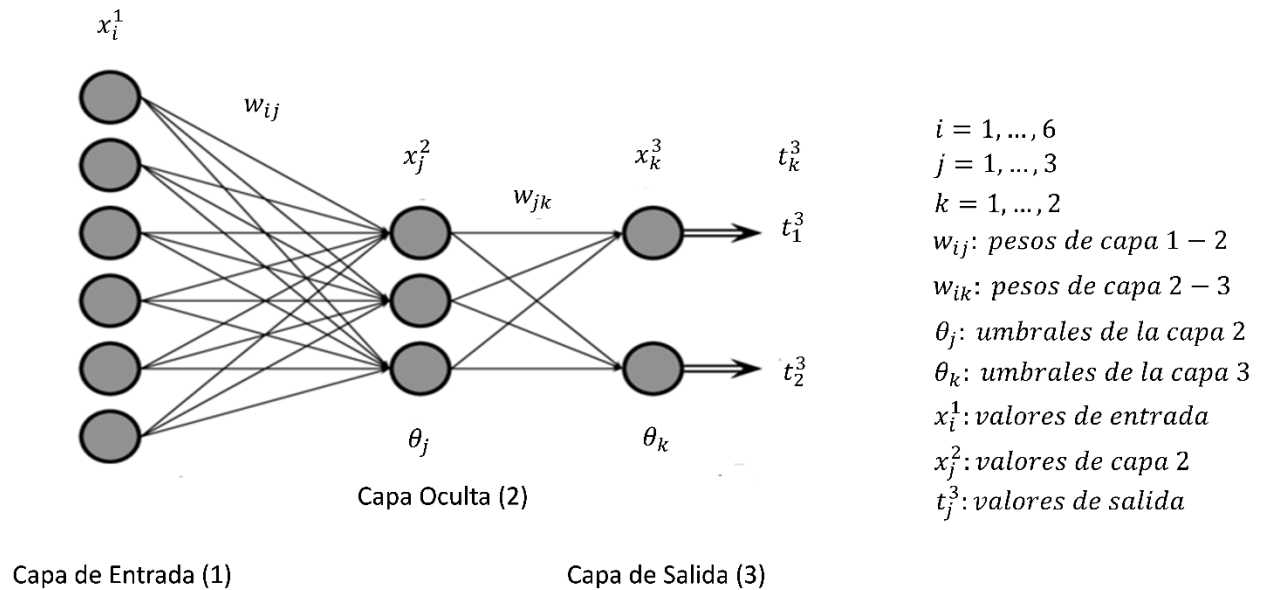


Figura A.19. Red neuronal formada por tres capas de neuronas artificiales (capa de entrada, capa oculta y capa de salida), y que constituye a lo que se conoce como el *Perceptrón Multicapa*.

A.5.2 Perceptrón Multicapa (MLP)

Una red neuronal está constituida por dos o más de estas neuronas. De hecho, puede tener varias capas de neuronas, cada una con decenas, cientos o miles de ellas. En la **Figura A.19** se muestra un arreglo típico que consta de tres capas: una capa de entrada, una intermedia u oculta, y otra de salida. i , j y k representan el número de neuronas en cada una de ellas, los valores w y θ son los pesos y umbrales respectivamente, mientras que las entradas son los valores x_i^1 , es decir, una instancia u observación conformada por seis muestras.

Cuando el modelo tiene más de dos capas se le conoce como *Perceptrón Multicapa*, y a diferencia del perceptrón simple es capaz de clasificar satisfactoriamente una amplia gama de problemas no linealmente separables. El proceso de variar los pesos y encontrar el mínimo absoluto de la función de error es más complicado pero hoy en día existen algoritmos eficientes como el de *Retropropagación*, que resuelven el problema eficientemente aprovechando la capacidad de cómputo que tienen los ordenadores actuales.

A.5.3 Redes Neuronales Convolucionales (CNNs)

Existen muchos tipos de redes que pueden resolver problemas de clasificación de diversas complejidades. El perceptrón multicapa, a pesar de haberse desarrollado hace más de 30 años, sigue teniendo una gran importancia hoy día, ya que aún en las redes más complejas suele aparecer como una parte fundamental de su estructura, conectando la etapa final del procesamiento con el clasificador que discrimina las categorías. Ese es el caso de las *Redes Neuronales de Convolución* (CNNs), un tipo de ANN que ha tenido mucho éxito en la clasificación y segmentación de imágenes, y últimamente también en el procesamiento de señales o series temporales.

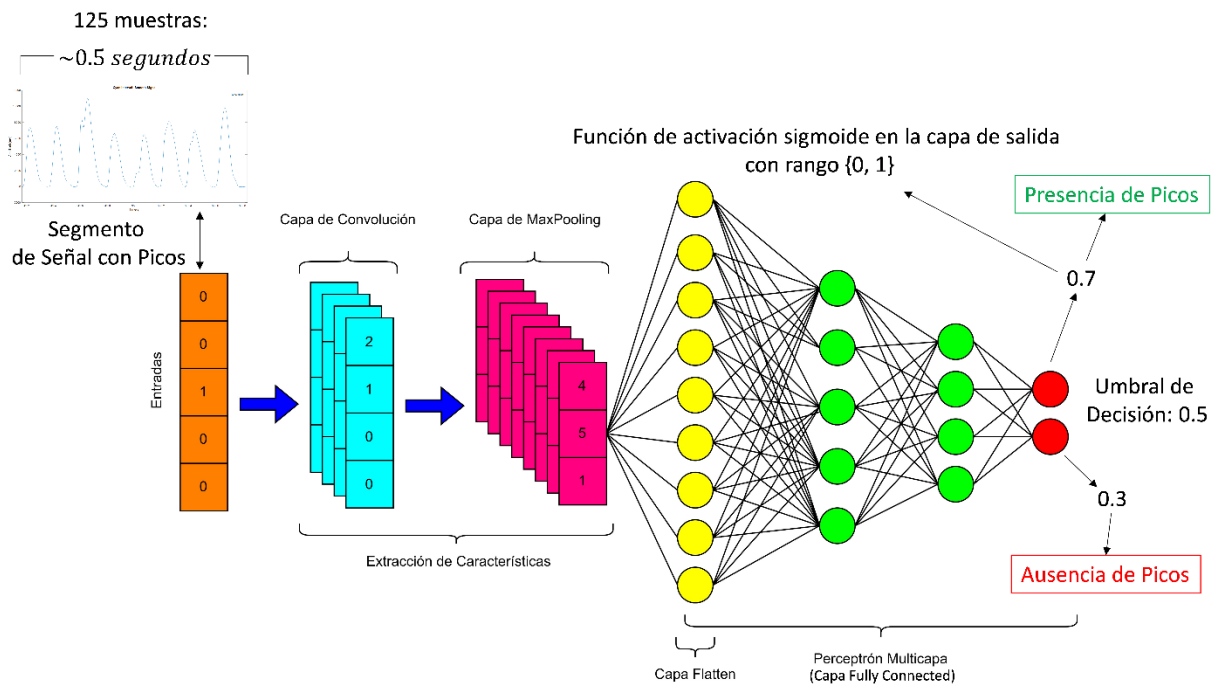


Figura A.20. Red neuronal convolucional que recibe segmentos de señales EMGs como entrada y las clasifica en dos categorías: con picos y sin picos.

Este tipo de redes son una extensión del perceptrón multicapa y añaden un sistema de capas adicionales, cuya estructura y operación está inspirada en el funcionamiento de los campos receptivos en la corteza visual primaria de un cerebro biológico. Por tanto, segmentan, filtran y aplican operaciones de reducción de muestreo para extraer las características principales de las

instancias (observaciones) y suministrarlas finalmente al perceptrón multicapa, donde también se ajustarán los pesos para identificarlas y clasificarlas.

Lo que se observa en la **Figura A.20** es una CNN con una estructura típica a la que se le suministra una serie de datos pertenecientes a un segmento de una serie temporal o señal unidimensional. Los segmentos (ventanas temporales de la señal) pueden ser de dos tipos, con picos y sin picos. Asumiendo que la red ha sido entrenada previamente con un gran número de segmentos de ambas categorías, muestra, ante una nueva instancia, la categoría a la que corresponde mediante la activación de la neurona respectiva en la capa de salida. Para que se active la neurona, su valor de entrada debe ser superior a 0,5, que es el umbral de decisión definido en este caso, y solo una de las dos neuronas puede activarse al mismo tiempo. La efectividad de la clasificación dependerá de la calidad y la cantidad de datos usados para su entrenamiento, así como del tipo y número de capas añadidas y los hiperparámetros utilizados en cada una. Para entrenar la red normalmente se utiliza el 80 % de los datos mientras que el 20 % restante se emplea en el proceso de validación. Se debe contar con una cantidad adecuada de datos para impedir el fenómeno de memorización de la red, el cual ocurre cuando ésta deja de clasificar adecuadamente al recibir datos diferentes a los que se usaron durante su entrenamiento. Cuando el modelo es capaz de clasificar adecuadamente datos que no fueron empleados en su entrenamiento, se dice que la red ha generalizado. El proceso de aprendizaje de la red consiste en modificar los pesos de las diferentes capas mediante algún algoritmo de optimización, e.g. *Adam*, para minimizar una función de pérdida o error. En problemas de clasificación la función de pérdida suele ser la *Entropía Cruzada*.

A.5.3.1 Principales Capas de una CNN

Una CNN se compone principalmente de capas convolucionales, max/average pooling, ReLus, dropout, fully connected y una capa de clasificación o regresión al final de la estructura. La arquitectura de la red puede cambiar de acuerdo al número de capas, su ordenamiento, y el valor de los hiperparámetros, de acuerdo con el tamaño y la complejidad de los datos. La convergencia y el grado de aprendizaje de la red también depende de los métodos de

optimización utilizados para cambiar los pesos. A continuación se describen brevemente cada una de las capas mencionadas.

A.5.3.1.1 Capa de Convolución

La convolución es una operación matemática que consiste en barrer un filtro de pesos a lo largo de las dimensiones de los datos de entrada, añadir un vector de umbrales, e integrar los resultados en neuronas artificiales para formar un mapa de características (**Figura A.21**). Este mapa contiene patrones detectados en los datos, aunque con una resolución más baja debido a la convergencia de la información. La convergencia se produce cuando la neurona artificial que conforma el mapa de características recibe como entrada la información de cada uno de los datos abarcados por el filtro sesgados por los pesos. Esto es similar a la recepción de señales presinápticas de distintas intensidades provenientes de varias neuronas al mismo tiempo en una neurona biológica. Los patrones capturados en un mapa de características pueden ser bordes verticales, horizontales, o inclinados, dependiendo de la disposición y magnitud de los pesos del filtro (**Figura A.22**). Al aplicar varios y diferentes tipos de filtros sobre los datos se genera un número equivalente de mapas de características con diferentes patrones reconocidos.

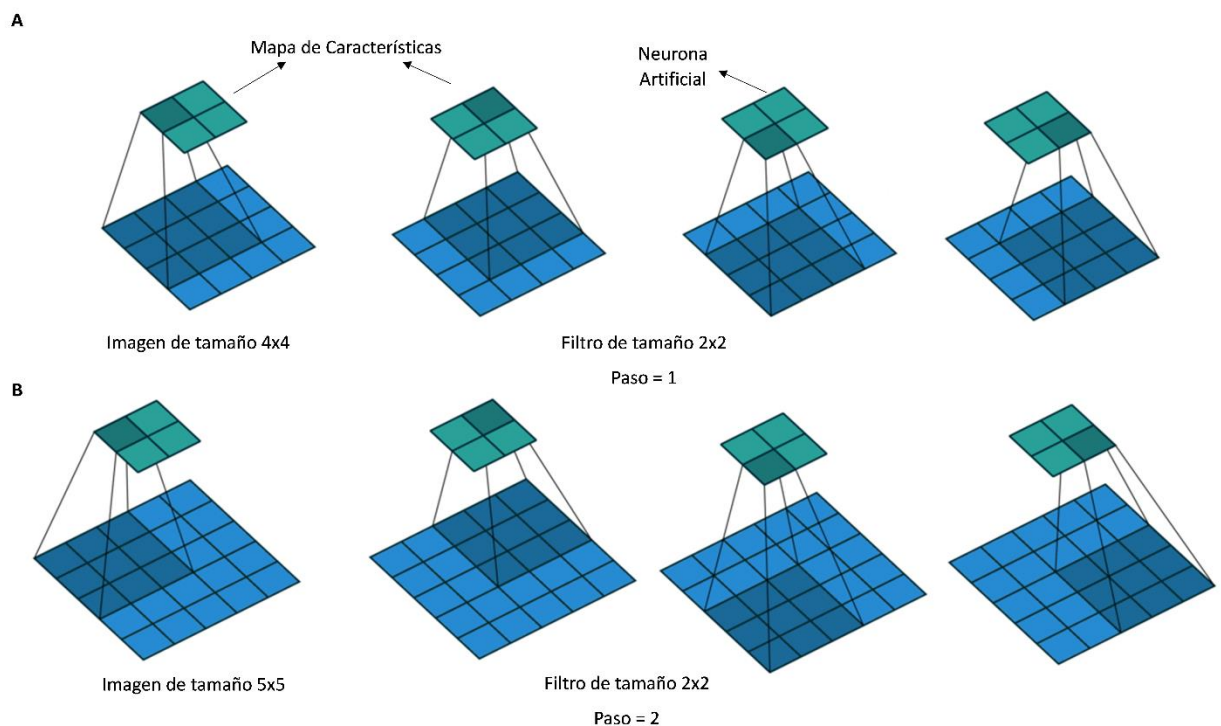


Figura A.21. Ilustración de una operación de *Convolución* bidimensional sobre una imagen de entrada aplicando un filtro con un paso determinado para formar un mapa de características. **A** – Filtro de tamaño 2x2 aplicado a una imagen de entrada de tamaño 4x4 con un paso de valor 1. **B** – Barrido de un filtro de tamaño 2x2 sobre una imagen de tamaño 5x5 con un paso de valor 2. Extraído de <https://www.mathworks.com/help/deeplearning/ug/layers-of-a-convolutional-neural-network.html>.

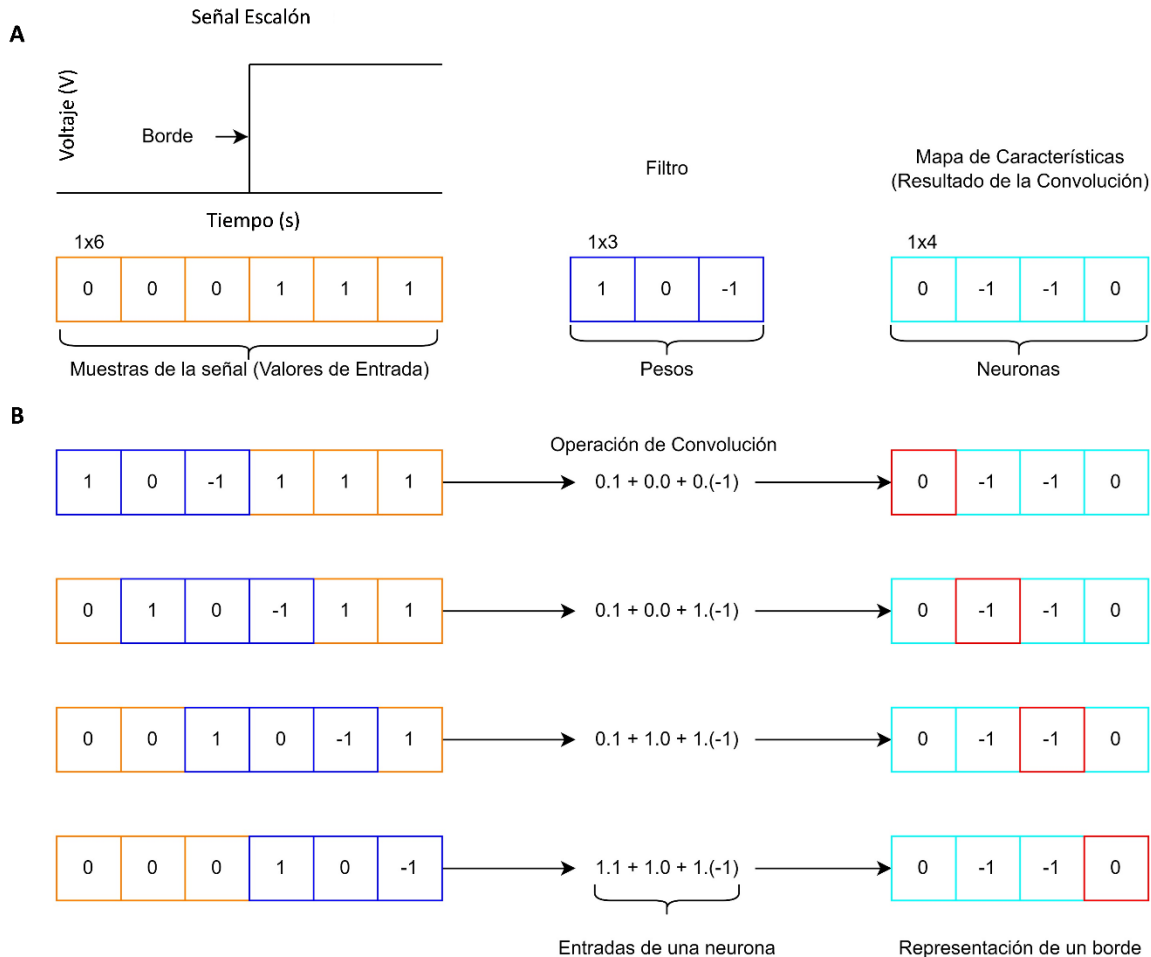
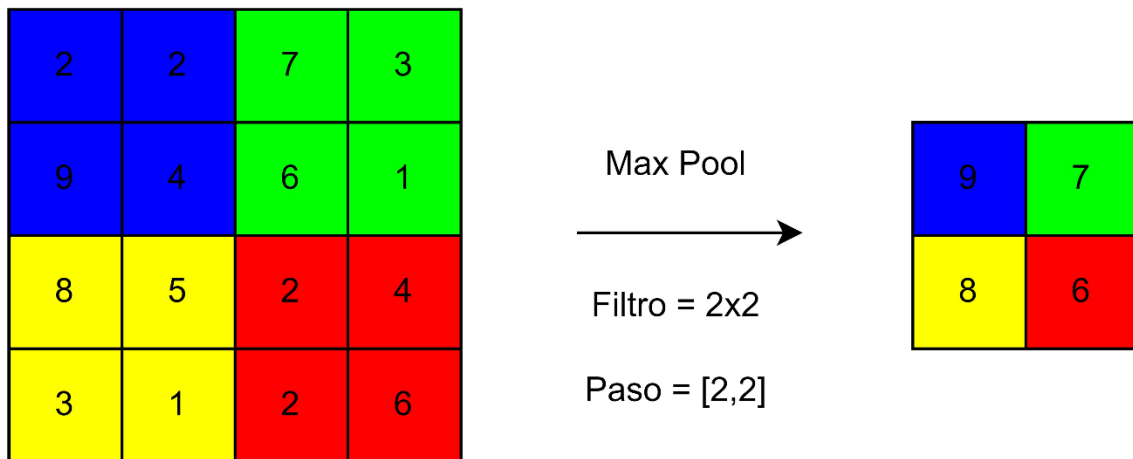


Figura A.22. Operación de *Convolución* unidimensional entre una señal escalón de voltaje vs. tiempo y un filtro de detección de bordes. **A** – Se muestra la señal escalón, el vector de entrada que representa la señal, el filtro para la detección del borde del escalón, y el resultado de la convolución que produce un mapa de características con una longitud menor al vector de entrada. **B** – Barrido del filtro sobre el vector de entrada durante la operación de convolución.

A.5.3.1.2 Capa de Max/Average Pooling

Esta capa normalmente se encuentra después de las capas de convolución para reducir el número de parámetros que deben aprender las capas siguientes, i.e, reducen el tamaño de los mapas de características o el número de neuronas que contienen. Esta operación puede ayudar a reducir la memorización o el sobreajuste de la red. El procedimiento consiste nuevamente en aplicar un filtro a los datos y obtener el valor máximo de aquellos que caen dentro del filtro, o el valor promedio en caso de que la capa sea de *Average Polling* (**Figura A.23**). Al no existir modificación de los pesos en esta capa no se produce ningún aprendizaje de parámetros.

A



B

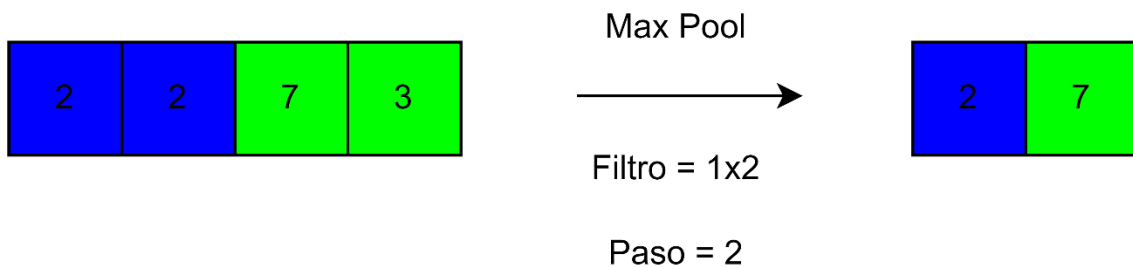


Figura A.23. Operación de *Max Pooling* en un mapa de características con un filtro y un paso determinado. **A** – Mapa de características bidimensional de tamaño 4x4, filtro de tamaño 2x2 y paso = 2. **B** – Mapa de características unidimensionales de tamaño 1x4, filtro = 2 y paso = 2.

A.5.3.1.3 Capa de Unidad Lineal Rectificada (ReLU)

En esta capa se aplica la función de activación ReLU a las neuronas de los mapas de características, la cual establece en cero las salidas con valores negativos y devuelve las positivas sin alteración (**Figura A.24**). Esta operación agiliza y mejora el aprendizaje con respecto a otras funciones de activación, por lo que es una de las más utilizadas en CNNs, MLPs y en general, en *Deep Learning*. La capa ReLU no modifica el tamaño de los datos de entrada.

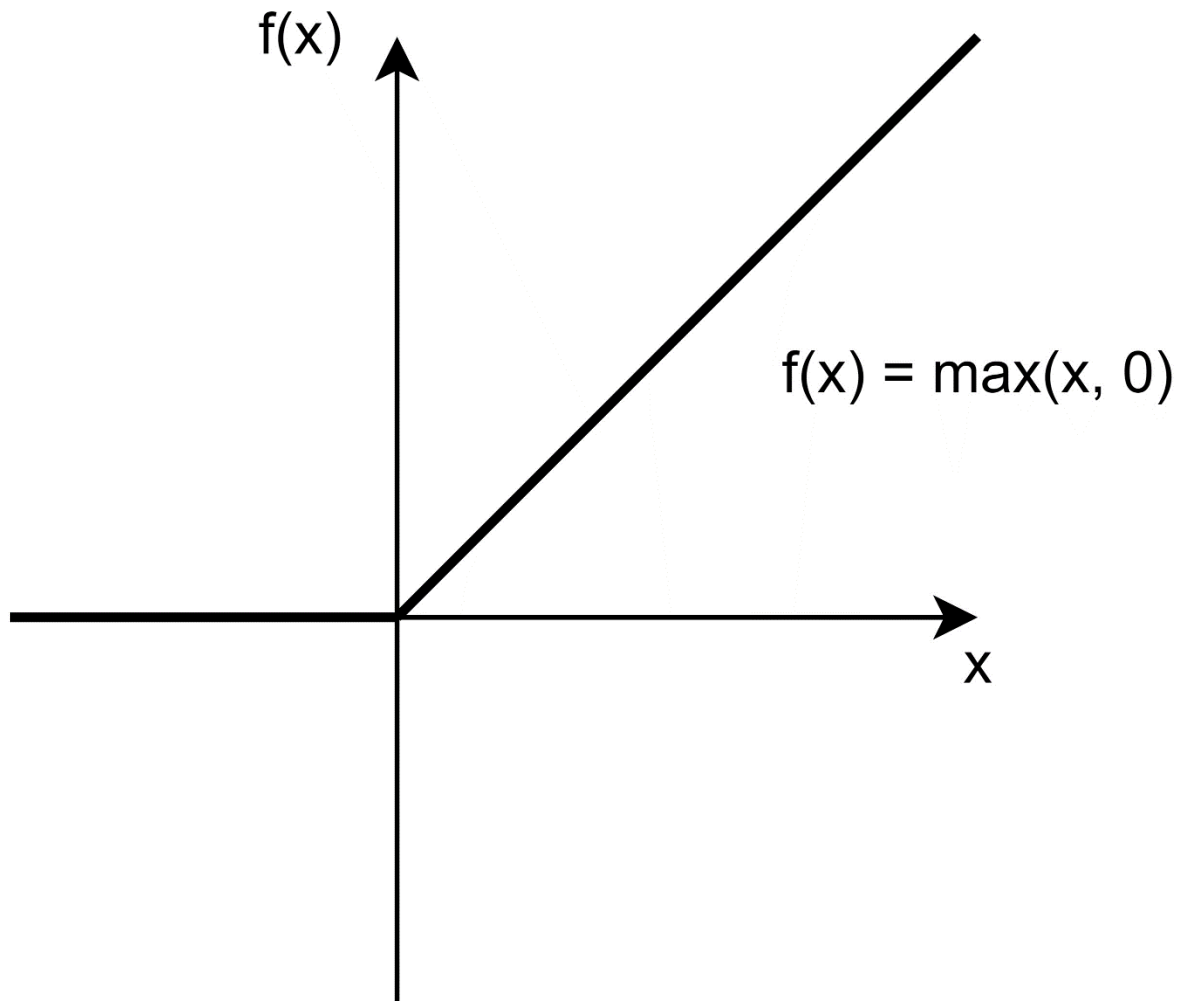


Figura A.24. Operación de umbral aplicada a cada uno de los elementos de entrada en una capa ReLU.

A.5.3.1.4 Capa Dropout

Esta es una capa de regularización en la que se apaga un número determinado de neuronas con determinada probabilidad (hiperparámetro) durante cada iteración del entrenamiento. Esto cambia la arquitectura de la red con cada iteración y ayuda a reducir el sobreajuste. (**Figura A.25**). Una mayor probabilidad apaga un mayor número de neuronas. Durante la predicción los datos que entran en la capa son iguales a los que salen, y al igual que en las capas de Max o Average Pooling, no existe aprendizaje o modificación de pesos.

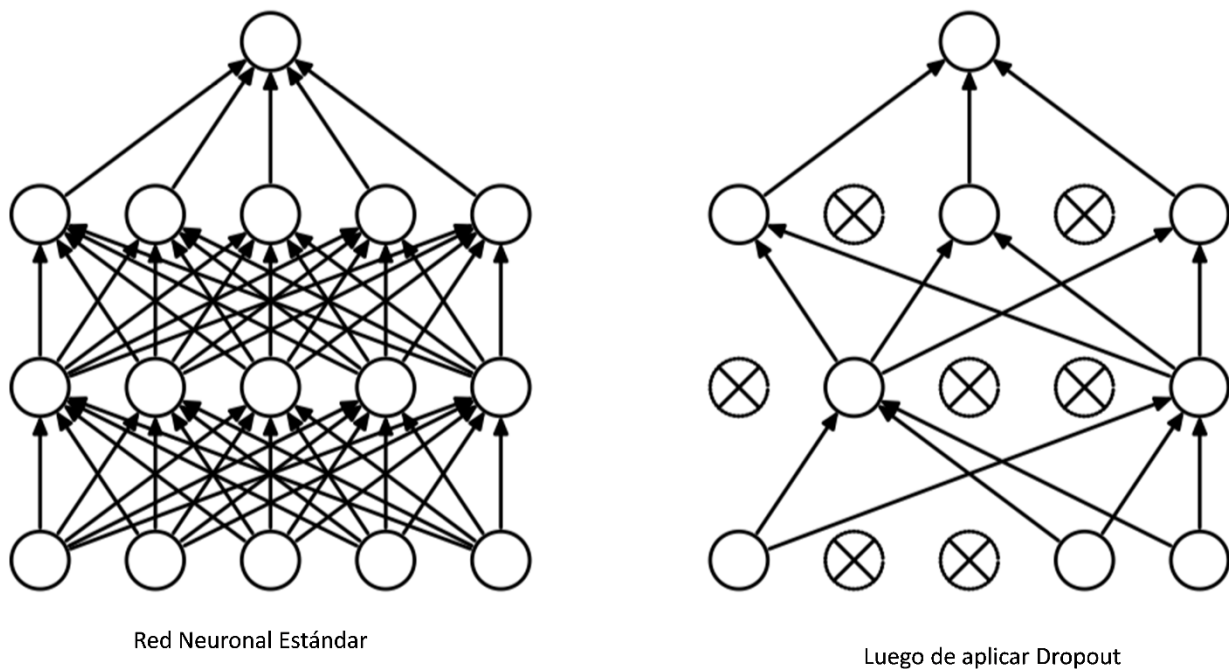


Figura A.25. Operación de regularización realizada sobre una red neuronal en la capa *Dropout*. Algunas neuronas se “apagan” de forma aleatoria con una cierta probabilidad (hiperparámetro) para incrementar la convergencia de la información y evitar la memorización de las características. Extraído de (Srivastava *et al.*, 2014)

A.5.3.1.5 Capa Flatten

En esta capa se establecen los batches de observaciones en un vector de una sola dimensión para que sirva de entrada a la capa *Fully Connected*.

A.5.3.1.6 Capa Fully Connected (FC)

En esta capa se aplica un perceptrón multicapa a los datos de entrada, i.e, se multiplican los valores de entrada por una matriz de pesos y luego se añade un vector de umbrales. Todas las neuronas de esta capa están conectadas a todas las neuronas de la capa previa. Si la capa previa es la capa Flatten, entonces se combinan todas las características aprendidas en capas anteriores para identificar los patrones más grandes. En una CNN es típico encontrar varias capas FC y con frecuencia, unas seguidas de otras. En problemas de clasificación suele existir una FC al final de la red que combina las características para clasificar las observaciones. Por esta razón el tamaño de su salida coincide con el número de clases del conjunto de datos.

A.5.3.1.7 Capa de Clasificación

La capa de clasificación es la que permite al modelo establecer la clase a la que pertenece una observación. Esto se hace mediante la aplicación de una regresión logística en caso de que existan solo dos clases. Para clasificar en un problema multiclase, se utiliza una regresión logística multinomial, también conocida como *Softmax*.

A.6 Redes Recurrentes

Las redes neuronales recurrentes (RNRs) son algoritmos usados para aprender patrones en datos secuenciales, como fragmentos de video o series temporales. Tienen la capacidad de utilizar la información obtenida durante el entrenamiento de datos previos en la secuencia para obtener las predicciones del dato actual y los siguientes. Esto se consigue comúnmente mediante la adición de transformaciones lineales con pesos ajustados en los pasos anteriores hasta alcanzar el paso final, que proporciona la salida del modelo. Por lo tanto, se puede pensar a una red recurrente estándar como una serie de operaciones de entrenamiento cíclicas donde el resultado de una depende de las anteriores. De forma más específica, para modelar los datos se aplica una red neuronal tradicional a cada elemento del conjunto de datos de entrada. Estos son comúnmente escalares, vectores (en el caso del IEVI), o matrices (imágenes), y la información de las capas ocultas de estas redes se adiciona en las capas ocultas de la red actual (**Figura A.26**). Esto le permite a la RNR memorizar los datos que son importantes para cada subred y olvidar aquellos que estas descartan con el uso de funciones de activación. Sin embargo, en la práctica

la memoria de las RNRs es de muy corto plazo, por lo que no son capaces de conectar la información del pasado con la del presente si el período en el que suceden es relativamente largo. Por esta razón surgió la necesidad de crear un modelo capaz de retener la información a largo plazo, y el resultado fueron las redes LSTM, que añaden una mayor complejidad respecto a las RNRs pero tienen un mejor desempeño.

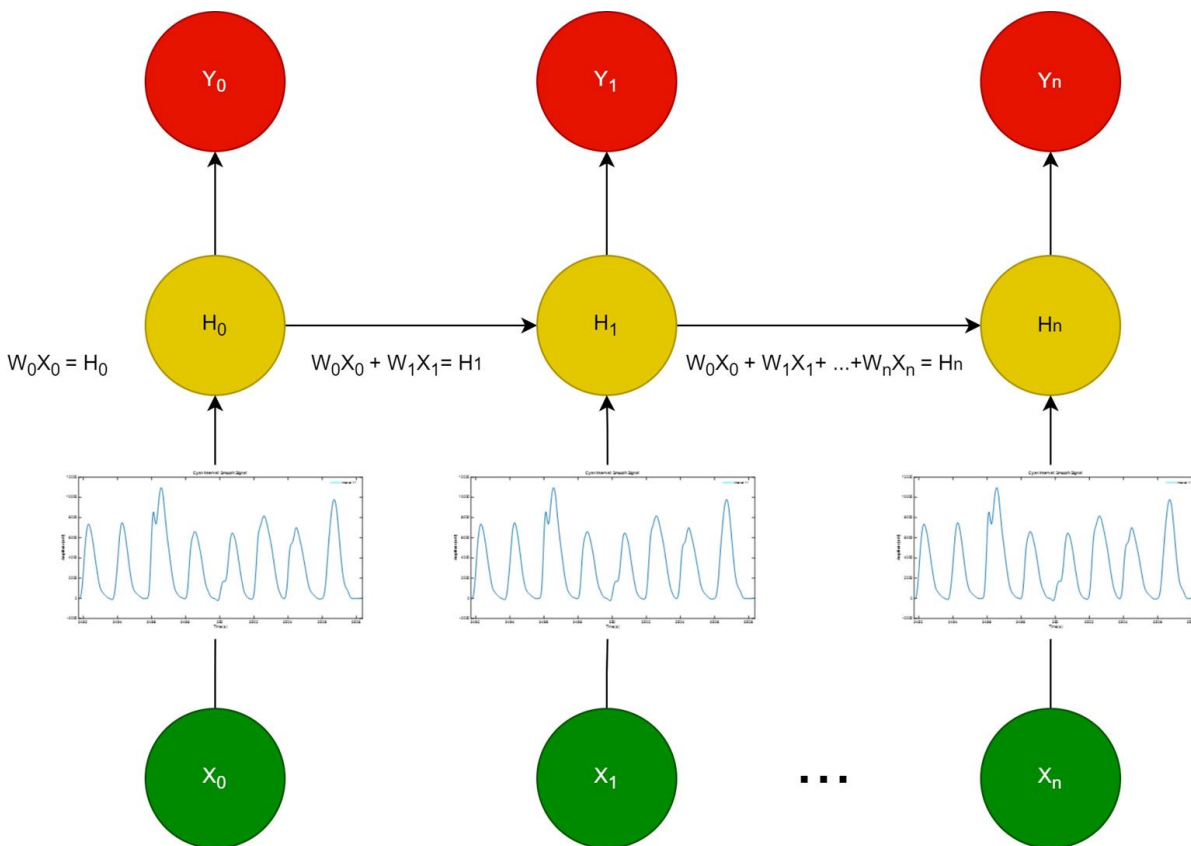


Figura A.26. Principio de funcionamiento de una red neuronal recurrente. Los segmentos de la señal representados por los vectores X_i son procesados mediante transformaciones lineales H_i para producir las salidas Y_i . En cada paso de la secuencia se transmite información de los pasos anteriores, lo cual robustece el proceso de aprendizaje del modelo.

A.6.1 LSTM

Las redes de memoria a largo y corto plazo (LSTM, por sus siglas en inglés de *long-term and short-term memory networks*), son una extensión de las redes recurrentes con la ventaja de

preservar información a largo plazo durante el entrenamiento. De forma similar a las RNRs en las que existe una red neuronal asociada a cada dato de entrada, las LSTMs también poseen elementos similares pero más complejos llamados células (**Figura A.27**), cuyo estado se modifica al descartar y agregar información mediante el uso de estructuras llamadas compuertas. Las compuertas se forman mediante la aplicación de una función de activación sigmoide seguida de una multiplicación puntual (F_t , I_t y G_t en la **Figura A.28**). La salida de un célula se transmite a la siguiente mediante dos vectores C y H , donde C se denomina estado de la célula y H estado oculto de la célula. El procesamiento de la información dentro de la célula tiene tres pasos.

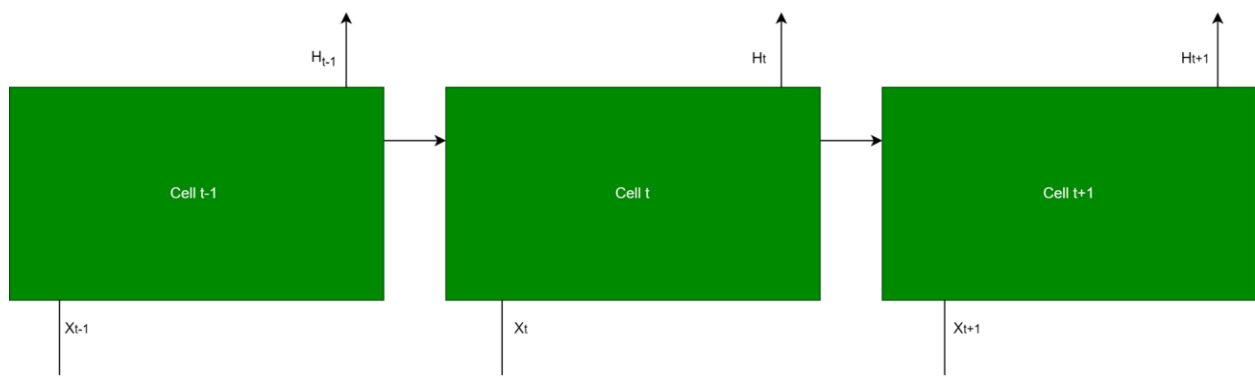


Figura A.27. Principio de funcionamiento de una red LSTM. Existe una célula para cada dato de entrada perteneciente a la secuencia donde se realizan operaciones que extraen la información relevante y la transmiten a la célula siguiente. Esto permite preservar la información a largo plazo.

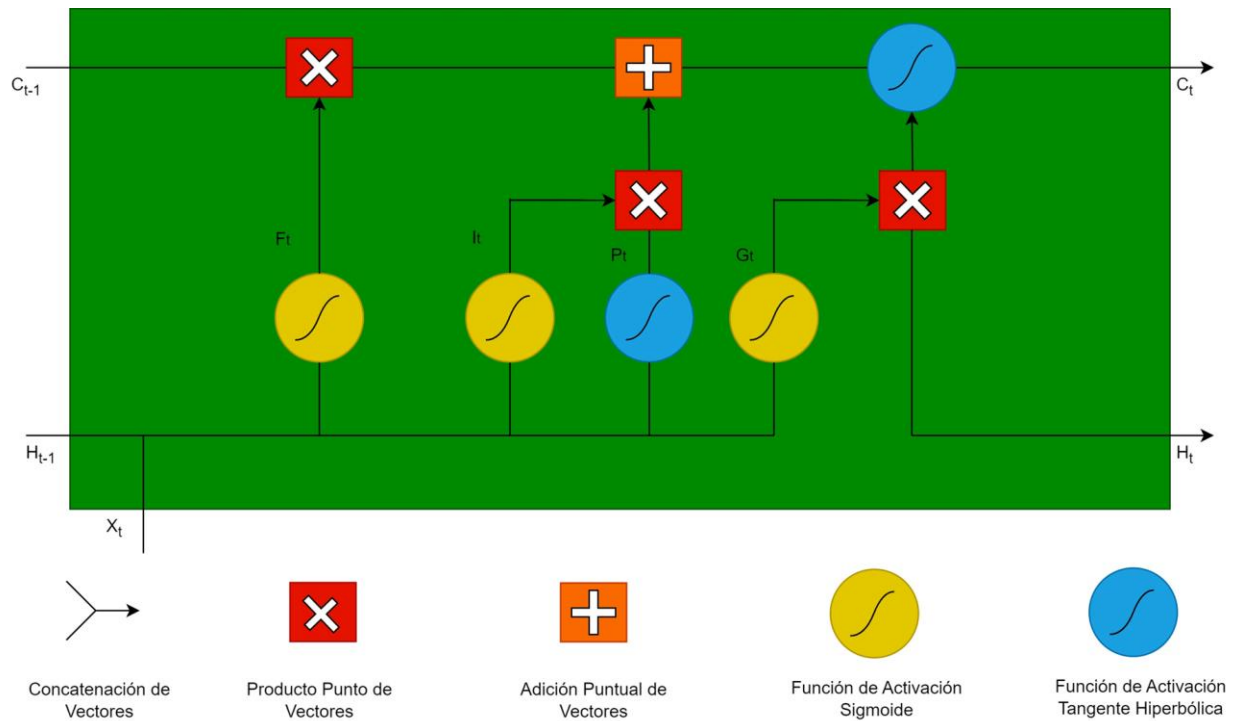


Figura A.28. Principio de funcionamiento de una célula perteneciente a una red LSTM. Los vectores C y H portan información proveniente de las células anteriores. El vector C se denomina estado de la célula y es modificado mediante una serie de operaciones matemáticas que culminan con la actualización de H . Las operaciones involucran la concatenación de vectores, aplicación de las funciones de activación sigmoide y tangente hiperbólica, así como productos y adiciones puntuales.

En el paso 1 se aplica la denominada compuerta de descarte sobre la concatenación de los vectores H_{t-1} y X_t y proporciona un número entre cero y uno para cada uno de sus elementos (**Figura A.29**). Un valor cercano a uno significa que el elemento prevalece, mientras que si su valor se acerca a cero se descarta. Es decir, el primer paso del proceso consiste en decidir qué información será descartada en el vector C . La operación matemática que describe la compuerta viene dada por:

$$F_t = \sigma(W_F[H_{t-1}, x_t] + b_F) \quad (39)$$

donde F_t es la salida entre cero y uno de la función de activación sigmoide sigma, W_F son los pesos aplicados al vector resultante de concatenar H_{t-1} y el vector de entradas X_t , mientras que b_F es un vector de pesos umbrales.

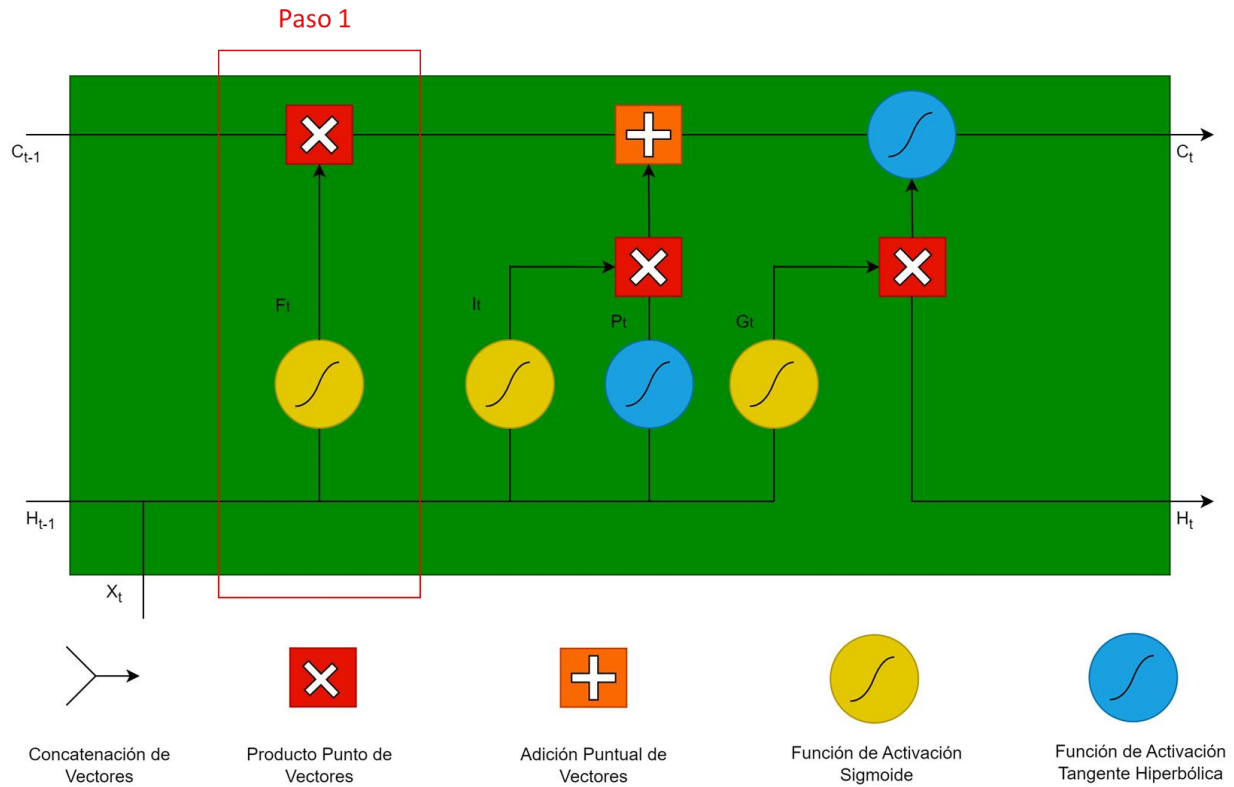


Figura A.29. Principio de funcionamiento de una célula perteneciente a una red LSTM. El primer paso consiste en aplicar la función sigmoide al vector resultante de concatenar H y X y realizar el producto punto con C . Este conjunto de operaciones recibe el nombre de “capa de compuerta de descarte”.

En el paso 2 se determina qué información nueva se almacena en el vector C . Primero se aplica una compuerta denominada “Compuerta de Entrada” a la concatenación de H_{t-1} y X_t . Al mismo tiempo se realiza otra operación similar con la diferencia de que se sustituye la función sigmoide por la tangente hiperbólica:

$$I_t = \sigma(W_I[H_{t-1}, x_t] + b_I) \quad (40)$$

$$P_t = \tanh(W_P[H_{t-1}, x_t] + b_P) \quad (41)$$

Luego estas dos operaciones se multiplican en forma puntual y se adicionan al resultado del paso 1 para actualizar el vector C (**Figura A.30**):

$$C_t = F_t * C_{t-1} + I_t * P_t \quad (42)$$

En el paso 3 se utiliza una compuerta nuevamente para decidir qué información estará en la salida de la célula. Se aplica la tangente hiperbólica sobre el vector de estado de la célula C y el resultado se multiplica por el de la compuerta para actualizar el vector de estado oculto, que junto a C representa la salida de la célula (**Figura A.31**):

$$G_t = \sigma(W_G[H_{t-1}, x_t] + b_G) \quad (43)$$

$$H_t = G_t * \tanh(C_t) \quad (44)$$

Posteriormente los cambios hechos en los vectores C y H se transmiten a la célula siguiente, lo que permite al algoritmo modelar los datos con un desempeño normalmente superior al de una RNR.

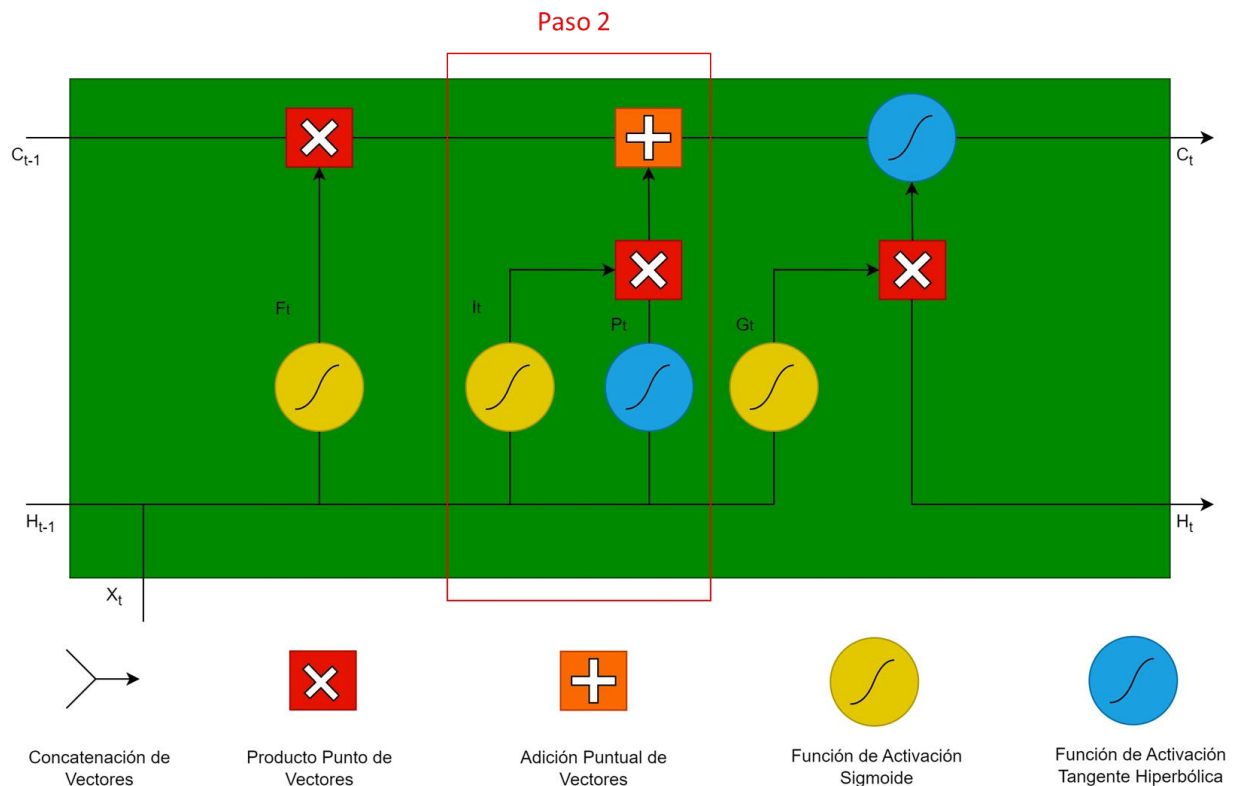


Figura A.30. Principio de funcionamiento de una célula perteneciente a una red LSTM. El segundo paso consiste en aplicar la función sigmoide al vector resultante de concatenar H y X , realizar la misma operación pero con la tangente hiperbólica, y luego sumar de forma puntual los vectores

resultantes para actualizar el vector C . Este conjunto de operaciones recibe el nombre de “capa de compuerta de entradas”.

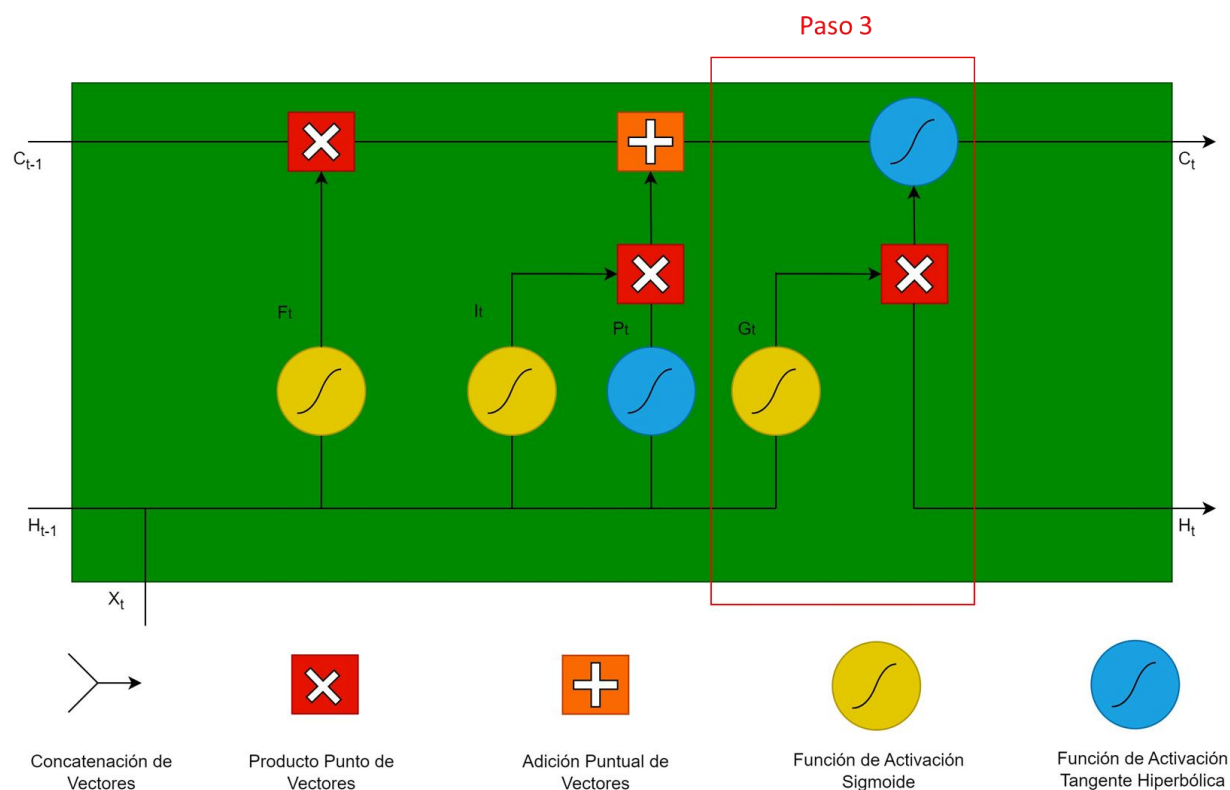


Figura A.31. Principio de funcionamiento de una célula perteneciente a una red LSTM. El tercer y último paso consiste en aplicar la función sigmoide al vector resultante de concatenar H y X , aplicar la tangente hiperbólica al vector C , y luego multiplicar de forma puntual los vectores resultantes para actualizar el vector H . Los nuevos vectores C y H conforman la salida de la célula.

APÉNDICE B

A continuación se presenta una tabla que describe los hiperparámetros utilizados por el Detector de Picos (PD) y el Calculador de Datos (DC).

Tabla B. 1 Hiperparámetros del PC y el DC.

Nombre del hiperparámetro	Abreviatura	Aplicación	Descripción
Factor de Amplitud	af*	PC, DC	Umbral de voltaje Preestablecido
Umbral de Pendiente	st*	PC, DC	Umbral de pendiente de la segunda derivada sobre la señal suavizada
Ancho de Suavizado	sw*	PC, DC	Número de puntos utilizados para suavizar la señal desde una muestra** dada.
Agrupamiento de Picos	pg*	PC, DC	Número de picos en el tope de los picos no suavizados que se utilizan para calcular su altura.
Tipo de Suavizado	sty*	PC, DC	Tipo de algoritmo utilizado para suavizar la señal EMG. Puede ser Rectangular (un paso de media móvil), Triangular (dos pasos de media móvil), Pseudo-Gausiano (tres pasos de media móvil) o Savitzky-Golay.

Factor de Varianza	wf	PC, DC	Establece un intervalo temporal alrededor del pico de voltaje y sustituye el valor asignado al pico suavizado por el punto máximo del mismo en la señal sin suavizar.
Factor Bajo de Varianza	lwf	PC, DC	Es igual al wf pero se aplica a picos que tienen menos de la tercera parte del tamaño del pico más alto de la FVI. Esto evita que se establezca como pico una muestra correspondiente a la ladera de algún pico aledaño más alto.
Factor Horizontal	hf	PC, DC	Establece un intervalo alrededor de una muestra identificada como pico y sustituye su altura por la del pico más alto presente en el intervalo.
Longitud de Ventana	wl	DC	Es la cantidad de muestras que componen una ventana temporal de la señal del EMG sobre la que se aplica la Transformada de Fourier.

Solapamiento	ovl	DC	Es el número de muestras solapadas entre dos ventanas temporales
--------------	-----	----	--

* Hiperparámetros basados en (O’Haver, 2023). Para más información ir a <https://terpconnect.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm#ipeak>

** Una muestra en este contexto (procesamiento de señales) se refiere a un valor puntual de voltaje medido por el osciloscopio.

APÉNDICE C

A continuación se muestran los hiperparámetros que dieron mejores resultados para cada uno de los tipos de modelos que fueron comparados en el **Capítulo 2.A**. Para iniciar la búsqueda de los hiperparámetros más adecuados correspondientes a los modelos de *Machine Learning* (Regresión Logística, SVM, ensambles de Random Forest y Gradient Boosting), se utilizó el método Tree-structured Parzen Estimators (TPE) mediante el uso de la librería Hyperopt: Distributed Asynchronous Hyper-parameter Optimization (<http://hyperopt.github.io/hyperopt/>), en el lenguaje de programación Python. Al encontrar una combinación prometedora de hiperparámetros se procedió con la técnica de Grid Search para realizar una búsqueda más refinada. Finalmente, la combinación más efectiva de hiperparámetros encontrada para cada tipo de modelo se muestra a continuación (solo se muestran los parámetros más relevantes). Para más detalles sobre hiperparámetros, procedimientos y estructura de los modelos, visitar <https://github.com/hecsalms/Model-Comparisons.git>

C.1 Modelos de Machine Learning

Tabla C. 1 Principales hiperparámetros de los modelos entrenados con datos *Raw* y *FFT*.

Modelo	Hiperparámetro	Librería	Raw	FFT
--------	----------------	----------	-----	-----

Regresión Logística (LR)	C	cuML	50.75	231.59
	max_iter		239	1310
Máquinas de Soporte Vectorial (SVM)	C	ThunderSVM	9.19	8.67
	gamma		0.02	0.15
Random Forest (RF)	max_depth	cuML	30	21
	n_estimators		127	108
Gradient Boosting (GB)	learning_rate	CatBoost	0.20	0.14
	max_depth		7	5
	n_estimators		250	260
Recurrent Neural Network (RNN)	optimizer__lr	Pytorch-Skorch	0.01	0.01
	batch_size		200	200
	max_epochs		80	80
Long Short-Term Neural Network (LSTM)	optimizer__lr	Pytorch-Skorch	0.001	0.01
	batch_size		200	200
	max_epochs		80	70

C.2 Redes Neuronales Convolucionales

En el caso de los modelos basados en CNNs, los hiperparámetros fueron encontrados mediante pruebas manuales. En las **Figuras C1 - C2** se muestra la arquitectura utilizada para la CNNt1 y una de las configuraciones empleadas para la secuencia CNNt2-LSTM, con los respectivos hiperparámetros para cada una de las capas que conforman ambos modelos. La red CNNt2, cuya arquitectura se describió en el capítulo 2A, también se utilizó en secuencias con los algoritmos de RF, GB y SVM (no se muestra). Estos modelos combinados se utilizaron únicamente en los períodos de prueba y no se incluyeron en el análisis final por razones ya mencionadas en el capítulo 2A. En los algoritmos RF, GB y SVM utilizados en secuencia con la CNNt2, se usaron hiperparámetros encontrados mediante búsqueda aleatoria con ajustes manuales.

```
Datos de Entrada: Raw, FFT  
  
Conv1d de 5x60 (in_chan= 1, ouput_channels = 5, kernel_size= 60, stride = 1, padding = 30, dilation = 1)  
ReLU()  
Conv1d de 10x30 (stride= 2, padding = 30)  
MaxPool1d (2)  
ReLU()  
Conv1d de 20x15 (stride = 2, padding = 30)  
MaxPool1d (2)  
ReLU()  
AdaptiveAveragePolling1d(29)  
Flatten(20x29)  
Tanh(FC de 580 a 300)  
Tanh(FC de 300 a 100)  
FC de 100 a 2
```

Figura C.1. Estructura de una red Neuronal Convolutacional de tipo I (CNNt1) que recibe datos de entrada Raw y FFT para clasificar dos tipos de patrones en los EMGs: con picos y sin picos.

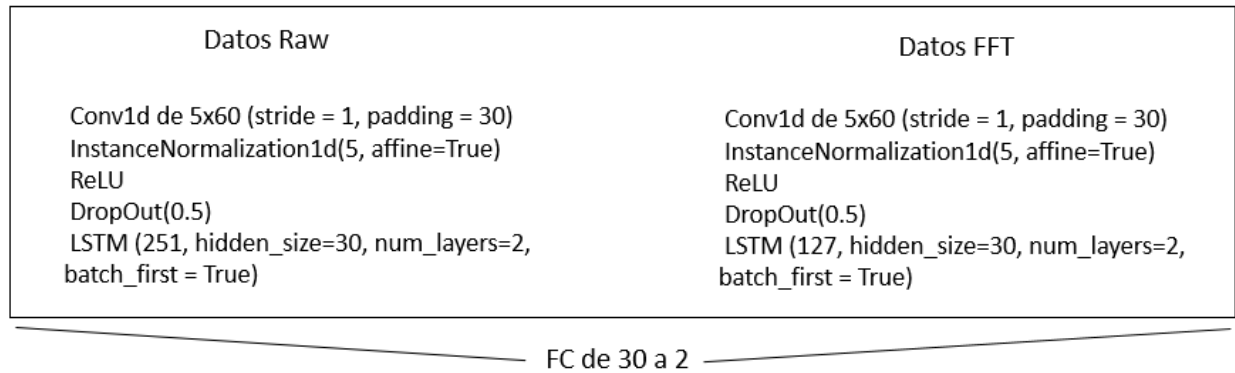


Figura C.2. Modelo CNNt2-LSTM como resultado de la combinación entre una red CNNt2 con una sola capa de convolución y una LSTM.

Director de tesis: Romina B. Barrozo

Co-director de tesis: Pablo E. Schilman

Consejero de Estudios: Gabriel Manrique

Tesista: Hector Luis Salas Morales