

La tesis comprende una serie de estudios destinados a mejorar los programas de selección genética para aumentar la eficiencia alimentaria en dos etapas de la vida productiva de los cerdos: durante la lactancia y el crecimiento. En diferentes estudios, correspondientes a los capítulos 2 a 5 de la tesis, se exploraron diversas metodologías y fuentes de información potencialmente útiles (incluyendo genética, genómica y fenómica) para lograr este objetivo.

La eficiencia alimentaria durante la lactación es un sistema complejo de rasgos fenotípicos interrelacionados. En el **Capítulo 2**, se emplearon modelos de ecuaciones estructurales para inferir las asociaciones causales entre la ingesta diaria de alimento, el balance de peso de la cerda, el aumento de peso de la camada, el balance del grosor de la grasa dorsal y el peso metabólico de la cerda. Se analizaron datos de 1.100 cerdas de un experimento de selección divergente para la eficiencia en cerdos en crecimiento. Se utilizó el algoritmo de causación inductiva para definir el modelo de ecuaciones estructurales. Los resultados indicaron que la ingesta diaria de alimento y el aumento diario de peso de la camada influyeron en el balance diario de peso de la cerda, con coeficientes estructurales medios posteriores de 0,12 y -0,03, respectivamente. Posteriormente, el balance diario de peso de la cerda impactó en el balance diario del grosor de la grasa dorsal y en el peso metabólico de la cerda, con coeficientes estructurales de 0,70 y -1,22, respectivamente. Se evaluaron múltiples estrategias de selección, incluidas aquellas que no dependen de los registros de ingesta de alimento. Además, utilizando la información de la estructura causal obtenida, se estimó la respuesta esperada a la selección en el caso de que se diera la misma cantidad de alimento a todas las cerdas. El objetivo de la selección fue siempre mejorar la productividad de la cerda sin perjudicar su condición corporal (es decir, reducir o mantener constante la ingesta diaria de alimento, aumentar o mantener constante el balance de peso de la cerda y aumentar el peso de la camada). Los resultados indicaron que todas las estrategias de selección propuestas llevaron a una respuesta similar, incluso aquellas que no dependían de los datos de ingesta de alimento, lo que haría la selección más fácil y económica. Sin embargo, parece difícil aumentar el peso de la camada mientras se reduce la ingesta de alimento sin perjudicar la condición corporal de la cerda.

En el **Capítulo 3**, se exploraron las ventajas de los métodos multi-output y stacking para mejorar la precisión de la predicción de la eficiencia alimentaria a partir del genotipo de cerdos en crecimiento. Se analizaron datos de 5.828 cerdos en crecimiento y 45.610 SNPs, implementando cuatro estrategias diferentes para predecir la eficiencia alimentaria. Dos métodos calcularon la eficiencia alimentaria de manera indirecta, ya sea a través de predicciones individuales de sus componentes (estrategia de múltiples salidas individuales) o mediante predicciones simultáneas (estrategia multi-output). Las otras dos estrategias predijeron directamente la eficiencia alimentaria, ya sea combinando predicciones individuales de sus componentes con datos de genotipo (estrategia

stacking) o utilizando solo los genotipos como predictores (estrategia de salida única). Esta última estrategia fue el punto de referencia. Se ajustaron modelos de random forest y support vector regression en todas las estrategias. Para evaluar la efectividad de cada estrategia, se implementó una validación cruzada anidada, que consistió en una validación cruzada externa de 10 pliegues y una validación cruzada interna de tres pliegues para el ajuste de hiperparámetros. Este proceso se repitió utilizando varios subconjuntos de variables predictoras con un número creciente (de 200 a 3.000) de los SNPs más informativos identificados con random forest. Para todos los subconjuntos de SNPs, el punto de referencia mostró el mejor rendimiento de la predicción. Con random forest como algoritmo de aprendizaje y los 1.000 SNPs más informativos como predictores, el promedio (SD) de los 10 valores obtenidos en los conjuntos de test fueron los siguientes: 0,23 (0,04) para la correlación de Spearman, 0,83 (0,04) para la pérdida cero-uno y 0,33 (0,03) para la pérdida de distancia de rango. En general, el estudio sugiere que incorporar información sobre los componentes predichos de la eficiencia alimentaria no mejora significativamente la calidad de la predicción en comparación con el método estándar de salida única, a menos que la calidad de predicción de uno o más componentes de la eficiencia alimentaria sea buena.

En cuanto a la fenotipado de alto rendimiento, se propuso el uso de los acelerómetros y las cámaras para obtener información sobre el comportamiento animal. En el **Capítulo 4**, dieciséis conejos fueron equipados con acelerómetros y grabados durante 2 semanas. Posteriormente, su comportamiento fue registrado manualmente y sincronizado con los datos del acelerómetro. Las técnicas de procesamiento de señales segmentaron las series temporales en ventanas de tiempo de diferentes tamaños (1, 3, 5, 7 y 9 segundos), y se calcularon un total de 41 características para cada ventana. Se entrenaron clasificadores de random forest, support vector machine y gradient boosting con las características más informativas para clasificar estados activos/inactivos. Se implementó una validación cruzada anidada, con una validación cruzada externa Leave-One-Animal-Out y una validación cruzada interna de tres pliegues para el ajuste de los hiperparámetros. Los resultados mostraron que los clasificadores funcionaron de manera similar. Con los hiperparámetros óptimos, utilizando una ventana de tiempo de 9 segundos y las 5 características más importantes, el modelo de random forest logró una precisión mediana de 1 y una recall mediana de 0,93, demostrando su alta confiabilidad en la clasificación de estados de inactividad en conejos en crecimiento.

Finalmente, en el **Capítulo 5**, se combinó un algoritmo de seguimiento por detección con un sistema de identificación por radiofrecuencia para abordar el desafío de rastrear doce cerdos en crecimiento alojados en grupo en un único corral. Entre ellos, tres cerdos tenían marcas naturales únicas en su pelaje, lo que facilitaba su distinción visual dentro del grupo. Los otros cerdos tenían patrones de color de pelaje similares o eran completamente blancos, lo que los

hacia visualmente indistinguibles entre sí. Se implementó la versión más reciente (v8) del algoritmo You Only Look Once para detección y el algoritmo BoT-SORT para el seguimiento. Se utilizaron un total de 3.600 imágenes para entrenar. Para los cerdos con colores de pelaje indistinguibles, se empleó el sistema de identificación por radiofrecuencia para identificar a los animales cada vez que entraban en la estación de alimentación. Este sistema de identificación por radiofrecuencia se vinculó con la trayectoria de cada cerdo, tanto hacia atrás como hacia adelante. Los resultados mostraron que los cerdos blancos fueron rastreados durante 59 minutos de un total de 166 minutos, los dos cerdos con marcas similares un promedio de 49 minutos y los cerdos con marcas únicas en el pelaje 152 minutos. En todos los casos, el tiempo de seguimiento asignado a cada cerdo coincidió con la verdad el 90% del tiempo o más. Este enfoque representa un paso significativo en la combinación de múltiples dispositivos para la identificación, seguimiento y trazabilidad de animales, especialmente cuando los animales no pueden distinguirse visualmente dentro de los grupos.

La tesis se completa con una contextualización del tema y una discusión general, destacando los hallazgos clave y sugiriendo direcciones para futuras investigaciones.