

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



“PREDICCIÓN DE PARÁMETROS ESTELARES Y ESTIMACIÓN DE INCERTIDUMBRE USANDO DEEP ENSEMBLE”

JOSÉ DOMINGO MONTECINOS ARISMENDI

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Dr. Raquel Pezoa, UTFSM
Profesor Guía: Dr. Michel Curé, UV
Profesor Correferente: Dr. Nicolás Rojas, UTFSM

Octubre - 2024

DEDICATORIA

A mis padres, hermanos y hermanas. A mis amigos. A todos aquellos sin los cuales no sería
quien soy.

AGRADECIMIENTOS

A mi querida familia y profesora Raquel Pezoa,

Quiero expresar mi sincero agradecimiento por todo el apoyo y la guía que me brindaron durante mi trabajo de título.

A mi familia, gracias por su comprensión y ánimo en los momentos difíciles; su confianza en mí fue fundamental para llegar hasta aquí.

A mi profesora Raquel Pezoa, le agradezco su paciencia, dedicación y valiosos consejos que hicieron posible este logro.

Este resultado es fruto del esfuerzo compartido. Gracias de corazón.

Con gratitud,

José Domingo.

RESUMEN

Resumen—Con la publicación constante de catálogos de estrellas, se hace necesario disponer de tecnologías que faciliten el tratamiento y análisis de datos a los astrónomos, agilizando así su trabajo y la toma de decisiones. Los parámetros estelares constituyen información crucial en el estudio de las estrellas, por lo que, contar con herramientas que entreguen estos parámetros es relevante. Recientemente, metodologías provenientes de la Inteligencia Artificial, han mostrado resultados favorables en el tratamiento de espectros estelares para la predicción de parámetros. En particular, las redes neuronales convolucionales tienen especial afinidad para resolver este tipo de problemas. En el presente trabajo se exponen los resultados de utilizar una metodología que emplea redes neuronales convolucionales, llamada *deep ensemble*, entrenando modelos a partir de espectros sintéticos de estrellas tipo B, para posteriormente efectuar una estimación de los parámetros T_{eff} , $\log g$, y $v \sin i$ de espectros observados del catálogo BeSOS. Los resultados arrojan una media de error para cada parámetro del 27 %, 9 % y 6 % respectivamente en este catálogo, además de conclusiones relevantes para el perfeccionamiento de esta y otras metodologías en investigaciones futuras.

Palabras Clave— Aprendizaje Profundo; Ensamble Profundo; Estimación de Incertidumbre; Predicción de Parámetros Estelares

ABSTRACT

Abstract— With the constant publication of star catalogs, it becomes necessary to have technologies that facilitate the processing and analysis of data for astronomers, thus speeding up their work and decision-making. Stellar parameters constitute crucial information in the study of stars, so having tools that deliver these parameters is relevant. Recently, methodologies from Artificial Intelligence have shown favorable results in the treatment of stellar spectra for the prediction of parameters. In particular, convolutional neural networks have a special affinity for solving this type of problem. This study presents the results of using a deep ensemble methodology, employing convolutional neural networks trained on synthetic spectra of type B stars, to estimate the parameters T_{eff} , $\log g$, and $v \sin i$ of observed spectra from the BeSOS catalog. The results demonstrate an average error of 27 %, 9 %, and 6 % for each parameter respectively in this catalog, along with significant conclusions for improving this and other methodologies in future research.

Keywords— Deep learning; Deep ensemble; Uncertainty Estimation; Stellar Parameters Prediction

GLOSARIO

DI: Departamento de Informática.

UTFSM: Universidad Técnica Federico Santa María.

DL: *Deep Learning* - Aprendizaje Profundo.

NN: *Neural Network* - Red Neuronal.

ANN: *Artificial Neural Network* - Red Neuronal Artificial.

CNN: *Convolutional Neural Network* - Red Neuronal Convolutacional.

DE: *Deep Ensemble* - Ensamble profundo.

MSE: *Mean Squared Error* Error Cuadrático Medio.

MAE: *Mean Absolute Error* - Error Absoluto Medio.

PDF: *Probability Density Function* - Función de Densidad de Probabilidad.

ÍNDICE DE CONTENIDOS

RESUMEN	IV
ABSTRACT	IV
GLOSARIO	VI
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABLAS	X
INTRODUCCIÓN	1
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	4
1.1 Contexto dentro de la Astronomía	4
1.2 El espectro de luz	6
1.3 Procesamiento y análisis computacional	8
1.4 La incertidumbre en la predicción	8
1.5 Objetivos	9
1.6 Contribuciones	10
CAPÍTULO 2: MARCO CONCEPTUAL	11
2.1 Fundamentos	12
2.1.1 Minimización del Riesgo Empírico	12
2.1.2 Función de Pérdida	13
2.1.3 Métricas de Desempeño	14
2.2 Red Neuronal Artificial	15
2.2.1 Red Neuronal Feed Forward	15
2.2.2 Red Neuronal Convolutiva	17
2.3 Cuantificación de la incertidumbre	20
2.4 Descomposición sesgo-varianza	21
2.5 Intercambio entre sesgo y varianza	22
2.6 Ensamble de modelos	23
2.7 Estado del Arte	25
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	27
3.1 Generación de espectros sintéticos	28
3.1.1 Tlusty	28
3.1.2 Synspec	28
3.1.3 Rotin	29
3.1.4 Línea de absorción del Helio	31
3.1.5 Configuración General para la Generación de Espectros	31
3.2 Implementación de Redes Neuronales	33
3.3 Red Neuronal Convolutiva Canónica	34

3.4	Encontrar el modelo de predicción óptimo con Autokeras	34
3.5	Ensamble de modelos	35
3.6	Predicción del ensamble sobre espectros observados	36
3.7	Verificaciones sobre imprecisiones en el modelo y datos	37
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN		38
4.1	Red canónica sobre datos no acotados	38
4.2	Red canónica sobre datos acotados	41
4.3	Optimización con Autokeras sobre datos acotados	42
4.4	Ensamble de modelos optimizados	44
4.5	Predicción sobre espectros observados	46
4.6	Discusión sobre el tratamiento de los datos y la predicción	47
4.6.1	CNN óptima en condiciones ideales	48
4.6.2	Ruido en en los datos	49
4.6.3	Discusión sobre la predicción en espectros observados de rango ex- tendido.	51
CAPÍTULO 5: CONCLUSIONES		58
ANEXOS		61
5.1	Anexo A	61
5.2	Anexo B	62
5.3	Anexo C	63
5.4	Anexo D	64
5.5	Anexo E	65
5.6	Anexo F	66
5.7	Anexo G	67
5.8	Anexo H	68
REFERENCIAS BIBLIOGRÁFICAS		69

ÍNDICE DE FIGURAS

1	Líneas de Fraunhofer.	6
2	Espectro de la estrella HD 172481.	7
3	Red Neuronal Artificial <i>feed forward</i>	16
4	Representación de una Red Neuronal Convolutiva típica	17
5	Ejemplo de Campo Visual y Kernel.	18
6	Representación del flujo de trabajo del ensamble de modelos.	24
7	Grilla de Parámetros: Estrellas de tipo B.	29
8	Espectro generado con Synspec.	30
9	Espectro generado con Synspec: Región de Interés.	30
10	Espectro Generado con Rotin.	31
11	Flujo de Generación de espectros sintéticos.	33
12	Ejemplo de espectro sintético con ruido añadido.	34
13	Arquitectura de CNN Canónica	35
14	Métricas de Entrenamiento, CNN Canónica	39
15	PDF's para predicción de CNN Canónica: Datos sin ruido añadido y de rango extendido.	40
16	Comparación para predicción de CNN Canónica: Datos sin ruido añadido y de rango extendido.	40
17	PDF's para predicción de CNN Canónica: Datos escalados, sin ruido añadido y de rango acotado.	41
18	Comparación para predicción de CNN Canónica: Datos de observados escalados y de rango acotado.	42
19	Representación de CNN Optimizada para datos acotados.	42
20	PDF's para predicción de CNN optimizada: Datos ruidosos y de rango acotado.	43
21	Comparación para predicción de CNN optimizada: Datos ruidosos y de rango acotado.	43
22	PDF's para predicción de ensamble de CNN optimizada (en verde) y datos de prueba (en rojo): Datos ruidosos y de rango acotado.	45
23	Comparación para predicción de ensamble de CNN optimizada: Datos ruidosos y de rango acotado.	45
24	Tendencia de la desviación estándar en datos de prueba para el ensamble de modelos.	46
25	Comparación para predicción de ensamble de CNN optimizada: Datos observados de rango acotado.	47
26	Representación de CNN Optimizada para datos de rango extendido.	48
27	PDF's para predicción de CNN optimizada: Datos sin ruido añadido y de rango extendido.	49
28	Comparación para predicción de CNN optimizada: Datos sin ruido añadido y de rango extendido.	49
29	PDF's para predicción de CNN optimizada: Datos con ruido y de rango extendido.	50

30	Comparación para predicción de CNN optimizada: Datos con ruido y de rango extendido.	50
31	Comparación para predicción de CNN optimizada: Datos observados de rango extendido.	51
32	Caso ideal de comparación entre espectro observado y sintéticos: HD157246.	52
33	Diferencias entre espectros observados y teóricos.	53
34	Desplazamiento de la línea de absorción	55
35	Ruido en la medición del espectro.	56

ÍNDICE DE TABLAS

1	Resumen de parámetros, Synspec y Rotin.	32
2	Separación del conjunto de datos.	33
3	Resumen de configuración de experimentos.	38
4	Métricas de desempeño, CNN canónica sobre datos sin ruido añadido y de rango extendido.	39
5	Métricas de desempeño, CNN canónica sobre datos sin ruido añadido y acotados.	41
6	Métricas de desempeño, CNN optimizada sobre datos sin ruido añadido y acotados.	42
7	Métricas de desempeño, ensamble de modelos optimizados sobre datos con ruido añadido y acotados.	44
8	Métricas de desempeño, ensamble de modelos optimizados sobre datos observados acotados.	47
9	Métricas de desempeño, CNN optimizada sobre datos sin ruido añadido y de rango extendido. Todas las métricas están calculadas sobre datos escalados.	48
10	Métricas de desempeño, CNN optimizada sobre datos con ruido adicionado y de rango extendido.	50
11	Tabla comparativa espectro observado vs sintético: HD157246.	52
12	Diferencias entre espectros observados y sintéticos. Tabla de parámetros.	54
13	Desplazamiento de la línea de absorción/ Tabla de parámetros.	56
14	Ruido en la medición del espectro.	57
15	Tabla resumen de resultados MAE.	57
16	Parámetros de entrenamiento Autokeras <i>bayesian</i>	61
17	Parámetros de entrenamiento Autokeras <i>greedy</i>	62
18	Predicción del ensamble de modelos sobre espectros de estrellas Be.	63
19	CNN Optimizada para datos acotados con ruido. Tabla de especificaciones.	64
20	Parámetros de entrenamiento Autokeras verificación	65
21	CNN Optimizada para datos de rango extendido y sin ruido. Tabla de especificaciones.	66
22	Parámetros para generación de espectros de Synspec.	67
23	Parámetros para generación de espectros de Rotin.	68

INTRODUCCIÓN

La espectroscopía es el área que estudia la interacción de los cuerpos con la radiación electromagnética, ya sea en forma de emisión, absorción o dispersión. Sus usos son cuantiosos en campos como la química, física, biología y astronomía [MacRae, 2019]. Esta última área es la que concierne al estudio realizado en esta memoria.

Los objetos celestes, específicamente las estrellas, emiten un espectro de radiación electromagnética al entorno, es decir un continuo de luz que, debido a su naturaleza ondulatoria, se puede descomponer en distintos rangos de longitud de onda o frecuencia. El espectro emitido por cada estrella es único y su estudio resulta particularmente interesante, ya que, es posible describir a la estrella a través de él. El espectro de emisión de una estrella se caracteriza por ser un continuo irregular de intensidad lumínica, existen declives que indican que la intensidad es muy baja y la estrella emite muy poca luz en ciertas frecuencias. El espectro de emisión nos dice mucho sobre la estrella, a través de él se puede determinar su temperatura, densidad, gravedad, distancia y luminosidad, además de los elementos presentes en su superficie y su abundancia. También, es posible establecer relaciones con la velocidad de rotación de la estrella además de otras características como temperatura efectiva, aceleración de gravedad, densidad, entre otros [Simón-Díaz, 2020].

En lo que concierne a este estudio, los parámetros estelares que nos interesa determinar, y cuya relevancia radica en que caracterizan a las estrellas de una manera particular, son: la temperatura efectiva (T_{eff}) la velocidad de rotación proyectada ($v \sin i$), y el logaritmo de la aceleración de gravedad ($\log g$).

Como se discutió previamente, las relaciones existentes entre el espectro y los parámetros resultan particularmente relevante en astronomía, dado que se puede obtener mucha información de las estrellas a través de su espectro. Esto por un lado resulta alentador desde la perspectiva del conocimiento que se puede generar sólo con un registro. Pero por otro lado, el universo es gigantesco y la cantidad de datos que se generarían si se quisiera tomar una muestra de cada estrella en el firmamento hace imposible su tratamiento y análisis desde el punto de vista de las limitadas capacidades humanas. Además, considerando la cantidad escasa de expertos en el tema, comparada con la cantidad de datos generados en el día a día y el tiempo que conllevaría estudiar datos a esta escala, describir adecuadamente estos cuerpos celestes resulta en una tarea monumental.

Con esta premisa es que se entiende la necesidad de contar con metodologías que trabajen estos datos de manera rápida y sistemática, de modo de otorgarle herramientas a los expertos que faciliten su trabajo y les permita dedicar más tiempo al análisis e investigación.

Métodos de tratamiento de datos originados desde la inteligencia artificial, en específico, las redes neuronales (NN) en el contexto del *Deep Learning* (DL) [LeCun *et al.*, 2015], han tomado relevancia en el último tiempo, debido a su éxito en tareas como análisis y producción de texto y sonido, robótica, clasificación de imágenes entre otros. Lo que nos interesa para

esta memoria es el análisis de espectros de emisión de estrellas a través de técnicas de inteligencia artificial, en particular, técnicas de aprendizaje automático y comprobar su efectividad en la predicción de los correspondientes parámetros estelares ya mencionados. Esto, además de generar una estimación de la incertidumbre asociada a cada predicción para poder contar con una noción de la incerteza que esta predicción tiene asociada.

En particular, dentro del de las NN, las redes neuronales convolucionales (CNN) tienen especial efectividad en el tratamiento de datos con estructura multidimensional o continuos. Por ende, una línea de emisión, que describe la intensidad lumínica emitida en una unidad de longitud de onda específica, en principio, tiene afinidad para ser tratada con esta metodología. La fortaleza de las CNN respecto de este tipo de datos es que no requieren que los datos sean excesivamente procesados, como sí necesitaban las técnicas tradicionales de *Machine Learning* (ML), donde era necesario que un experto seleccionara características importantes y las ordenara para ser trabajadas por una de estas técnicas. En el caso de las CNN, el aprendizaje de los datos de entrada está enteramente entregado a la “máquina” (entendiendo “máquina” como una una unidad de procesamiento capaz de realizar una predicción), de manera que esta puede realizar su propia representación de los datos, por complejos que sean, lo que le da a esta herramienta una poderosa capacidad de adaptación e interpretación. Es por este motivo que este tipo de arquitectura de NN es la principal herramienta en la predicción de parámetros estelares en este trabajo, dada su probada efectividad en diversos ámbitos [Li *et al.*, 2021] y también para este en particular [Zhao *et al.*, 2022].

Por otro lado, ya existen metodologías dentro del *deep learning* para la estimación de parámetros, sin embargo, estos métodos tienen excesiva confianza en mecanismos deterministas y no consideran la incertidumbre asociada la predicción. Es por eso que se hace necesario abordar la temática de la predicción de parámetros con una metodología más robusta, que cuantifique la incertidumbre de la medición realizada por un modelo, y genera una noción de incerteza en la predicción. Este tipo de aproximación resultará en una metodología más completa de trabajo, al hacer consciente a un experto que trabaje con ella del error correspondiente a cada predicción, lo que facilitará el análisis de los resultados y posteriores conclusiones.

En este trabajo se hará revisión y comparativa de un método tradicional de predicción de parámetros estelares con respecto a un método capaz de cuantificar incertidumbre, conocido como ensamble de modelos [Hastie *et al.*, 2009], con el objetivo de esclarecer la certeza con la que una máquina es capaz de realizar una predicción, además de dar luces en el campo de los métodos de ensamble, entregando resultados reproducibles y comparables con aquellos obtenidos con métodos del estado del arte de la determinación de parámetros estelares.

El trabajo se desarrollará bajo el análisis de espectros de modelos sintéticos de estrellas tipo B según el esquema de clasificación de Harvard, que van de los 10.000[K] a los 30.000[K], la segunda clasificación mas caliente, y se caracterizan por ser bastante luminosas, supergigantes según el esquema de clasificación de Yerkes, y presentar líneas de absorción de Helio neutral. También se verificará la efectividad de los métodos de predicción de parámetros con espectros observados de estrellas con clasificación tipo Be, estrellas tipo B enanas y que pertenecen a la secuencia principal [Allen, 2005].

El desarrollo del trabajo comprende cuatro capítulos. En el Capítulo 1, se establece una definición del problema junto con los objetivos y contribuciones asociadas. En Capítulo 2, se establece un marco de trabajo conceptual, donde se describen los fundamentos teóricos en los que se basa esta investigación. En el Capítulo 3, se da a conocer la propuesta de solución, detallando la metodología, además de los pasos del trabajo a desarrollar. Posteriormente, en Capítulo 4, se dará cuenta de la validación de dicha propuesta, entregando los resultados de la experimentación y junto con su respectivo análisis. Por último, en el Capítulo 5, se establecen conclusiones en base a los resultados arrojados.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

Actualmente, el estudio de las estrellas está íntimamente relacionado con el análisis de su espectro. Los parámetros estelares derivados de este análisis nos permiten dilucidar diversas características de las estrellas. Sin embargo, este proceso conlleva una metodología y labor que no son menores, incluso si el proceso es llevado por un experto. A continuación se introducirá la problemática que se desea tratar, y la metodología y objetivos que se proponen para abordarla.

1.1. Contexto dentro de la Astronomía

La determinación de los parámetros estelares T_{eff} , $\log g$, $v \sin i$ (temperatura efectiva, logaritmo de la gravedad y velocidad de rotación) tiene una gran relevancia dentro de la astronomía ya que, junto con la composición química (no revisada en este trabajo), permiten la caracterización de una estrella.

Generalmente las estrellas se describen en términos de su masa, luminosidad, radio, composición química, campo magnético, momento angular, entre otros parámetros. Varios de estos parámetros no son directamente observables, por lo tanto, se hace necesario utilizar otros métodos para inferirlos. Por ejemplo, los parámetros T_{eff} y $\log g$ no solo entregan información de la atmósfera de una estrella sino que también permiten determinar otros parámetros a través de relaciones, como la luminosidad en la relación de radiación de cuerpo negro, donde la T_{eff} es relevante, o también la relación de la aceleración de gravedad a partir de la masa y el radio de una estrella [Young *et al.*, 2008], siendo en ambos casos, el radio de la estrella un valor posible a determinar en dichas relaciones.

Por otra parte, la velocidad de rotación también cumple un rol fundamental en la caracterización de una estrella. En lo que respecta a este estudio en particular, resulta relevante tener en consideración el efecto Doppler [Young *et al.*, 2008] producido por la rotación, dado que, desde el punto de vista de un observador estático, el espectro luminoso que emana desde el lado de la estrella que rota en dirección al observador, tendrá una frecuencia menor en relación al espectro real, y el espectro obtenido desde el lado que rota en dirección opuesta, tendrá una frecuencia mayor. Esto, además de tener en cuenta que la velocidad de rotación es mayor en el ecuador que en el resto del cuerpo, por lo que habrán diferencias respecto al espectro obtenido según la zona observada. Por lo tanto, contar con la velocidad de rotación permite realizar las correcciones necesarias para generar inferencias ajustadas. Y en lo que respecta a otras características generales, la rotación misma de la estrella tiene otros efectos adicionales en las estrellas, como por ejemplo, el achatamiento de la misma y el oscurecimiento gravitatorio producto de la variación de temperatura en función de la latitud estelar generada por la rotación, la velocidad de los vientos solares, entre otros [Simón-Díaz, 2020].

En adición lo anterior, de acuerdo a [Simón-Díaz, 2020], tareas que requieren fuertemente de la parametrización de las estrellas incluyen, y no se limitan a:

- **Clasificación estelar**, como requisito preliminar a un análisis espectral mas detallado.
- **Modelos atmosféricos estelares** están determinados por características atmosféricas como la temperatura efectiva, aceleración de gravedad, abundancia de elementos, entre otros.
- Comprender la **estructura y evolución de la Vía Láctea** requiere un conocimiento profundo en cuanto a propiedades fundamentales de las estrellas, tales como masa, edad, movimiento y abundancia química de poblaciones estelares.
- Análisis de la **evolución química** de las galaxias permite establecer relaciones e historia entre estrellas, además de tantear modelos respecto de la generación de elementos en estrellas a través de reacciones termonucleares y supernovas.

Tomando en cuenta lo anterior, es posible entender la importancia que tiene la estimación de los parámetros estelares. Sin embargo, la labor que esta tarea requiere a día de hoy termina siendo un trabajo no menor, ya que se requiere llevar a cabo de diversos pasos, entre ellos los siguientes:

- Selección de los espectros estelares a analizar según la clasificación estelar.
- Selección de los conjuntos de líneas espectrales adecuadas a analizar.
- Determinar de los parámetros causantes de perturbaciones en las líneas espectrales
- Estimación de parámetros preliminares de la estrella necesarios para correcciones.
- Selección de los modelos de estrellas con parámetros bien conocidos para comparación con los espectros a analizar.
- Generación de una grilla de comparación de los parámetros que se desean determinar.
- Estimación de los parámetros estelares accesibles según las comparaciones previamente establecidas.
- Determinación de abundancias dentro de los elementos que se puedan analizar desde sus respectivas espectrales.

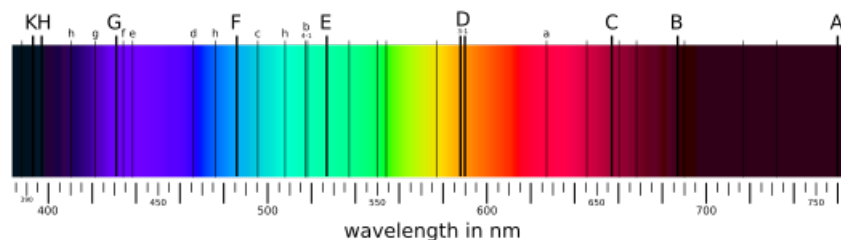


Figura 1: Líneas de Fraunhofer. En la figura se observa el espectro de emisión de nuestro Sol. Los elementos en su superficie producen las líneas espectrales observadas y corresponden a los siguientes: A y B: O_2 , C: H_α , D: Na, E: Fe, F: H_β , G: Ca/Fe, K y H: Ca^+ . Fuente: wikipedia.org.

1.2. El espectro de luz

La base de investigación de esta memoria es el espectro de una estrella. Para tener mayor claridad sobre qué es el espectro, es que se definen los espectros de emisión y absorción de acuerdo a [Young *et al.*, 2008]:

- **Emisión:** Corresponde al fenómeno observado al descomponer el haz de luz emitido por un material cuando este se encuentra a una alta temperatura. En este caso el material a alta temperatura se constituye por el propio contenido de la estrella, que emite luz en un continuo de todas las longitudes de onda.
- **Absorción:** Dentro del contexto de espectros estelares, corresponde a aquellas longitudes de onda en que se observa luz en baja intensidad debido a la absorción de los elementos presentes en la superficie de la estrella, lo que se percibirá como un declive en la intensidad lumínica medida.

En la Figura 1 es posible observar un ejemplo para ambos tipos de fenómeno. A partir de la idea de líneas observadas en el espectro, es que se desprende el concepto de “línea espectral”. Este concepto describe con precisión una línea observada en una longitud de onda específica dentro del espectro de una estrella.

Comprendiendo los fenómenos de emisión y absorción, se define el perfil espectral P-Cygni de acuerdo a [Templeton, 2009], de relevancia particular en astronomía:

- **P-Cygni:** Corresponde al fenómeno de observar tanto emisión como absorción en un rango de longitud de onda muy cercano. La luz emitida por una estrella entra dentro del fenómeno de emisión, sin embargo, puede ser que la luz sea absorbida por gases que rodean a la estrella en una cierta línea espectral o una muy cercana, por tanto se observarán fenómenos tanto de absorción como emisión.

En base a estos conceptos es que se pueden analizar los espectros estelares. La medición típica se realiza mediante un espectrógrafo [Gutiérrez Albores y Luna Aguilar, 2004] y se lee

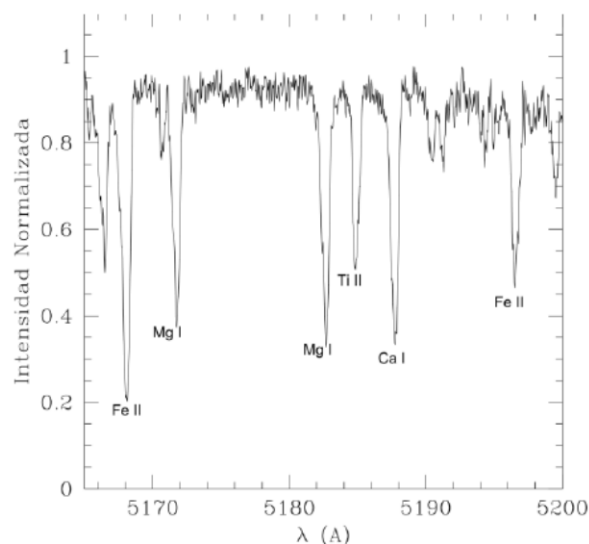


Figura 2: Espectro de la estrella HD 172481. La gráfica muestra intensidad lumínica percibida para cada longitud de onda. Se pueden apreciar las distintas depresiones en intensidad, correspondientes con las líneas de absorción de los elementos abundantes en la superficie de la estrella. Cada depresión tiene una profundidad y ancho diferente dependiendo de la abundancia del elemento y otros factores [Levy *et al.*, 2017].

en relación a la intensidad lumínica percibida respecto de la longitud de onda. En la Figura 2, se puede observar el espectro de la estrella HD 172481. También se puede apreciar que el espectro no es del todo continuo, si no que hay valles, es decir, zonas en las que la intensidad de luz es baja y existe menor emisión de luz en esa longitud de onda. Cada uno de esos valles se corresponde con una línea espectral de algún elemento presente en la estrella y se produce debido a la luz que absorbe dicho elemento. El conjunto de líneas espectrales determina lo que se conoce como el espectro de absorción de la estrella. El análisis comparativo de espectros de absorción de diferentes elementos con el de la estrella permite establecer cuáles elementos están presentes en la superficie de esta. Por otra parte, también es posible inferir más características acerca de la estrella y que resultan más pertinentes a este estudio. La profundidad de la línea espectral en una longitud de onda específica puede entregar detalles acerca de la abundancia del elemento asociado a esa línea. La presencia de ciertos conjuntos específicos de líneas espectrales está estrechamente relacionado a la determinación de la aceleración de gravedad y la temperatura efectiva, esto debido a la generación de iones específicos a partir de los elementos presente en la superficie de la estrella. Por otra parte, el ancho de las diferentes líneas espectrales (que también está relacionado a la determinación de aceleración de gravedad y la temperatura efectiva) permite estimar la velocidad de rotación de la estrella debido al efecto Doppler que esta última genera sobre el rango de la longitud de onda observada en una línea espectral. Estos y otros motivos verifican al espectro de absorción como una pieza clave en el estudio y caracterización de los astros.

1.3. Procesamiento y análisis computacional

Un espectrógrafo es una herramienta extremadamente útil para el estudio de estrellas. Permite obtener datos y hacer inferencias a través de espectros medidos, además de generar una gran cantidad de datos en poco tiempo. Sin embargo, también existe el problema de que una cantidad muy grande de datos es difícil de tratar con métodos convencionales, dada la acuciosa labor que se requiere.

La determinación de los parámetros estelares, además, es una tarea que requiere ser realizada con una buena cuota de confianza. La selección de: el conjunto de estrellas cuyos espectros se desea analizar, las líneas espectrales y los modelos de comparación, son pasos que requieren de personal experto para ser ejecutados de manera óptima, y que además, dada su naturaleza basada en la experiencia, resultan en labores difíciles de especificar en un algoritmo, lo que dificulta su sistematización.

Es por ello que surge la necesidad de contar con un mecanismo de apoyo que pueda sistematizar el proceso, incorporando en la medida de lo posible el conocimiento de los expertos y generar una aproximación basada en la experiencia. Dentro de esto, las redes neuronales resultan en una herramienta atractiva para la resolución del problema, dada su capacidad para aprender de la experiencia, y generar soluciones en un corto plazo.

1.4. La incertidumbre en la predicción

Muchas aplicaciones prácticas se han desarrollado de manera efectiva en el campo del *deep learning* (DL) al día de hoy. En diversas áreas, las más famosas actualmente pueden verse en tareas como el procesamiento de texto, generación de imágenes, análisis de señales continuas, entre otros. Su potencial radica en su capacidad para resolver problemas que no somos capaces de resolver por medio de una detallada lista de instrucciones. Lo que convierte a las herramientas generadas a partir del DL en soluciones atractivas en contextos con este problema.

Sin embargo, una temática constante relacionada al uso de inteligencia artificial, es tratarla como solución de tipo “caja negra”. Es decir, una máquina cuya naturaleza hace difícil describir con exactitud su método de análisis y procesamiento, por lo que se la usa sin mayor conocimiento de su mecanismo interno. Algunas metodologías que permiten mitigar dicho problema consisten en realizar una interpretación de los resultados que arroja la máquina, o establecer una trazabilidad de los pasos que se siguieron antes de establecer una decisión, metodologías que se enmarcan en el contexto de la IA explicable [Arrieta *et al.*, 2020] [Minh *et al.*, 2022].

Como consecuencia del tratamiento de caja negra, es que frecuentemente se tiene excesiva confianza respecto a los resultados que un modelo pueda entregar. Ya que no se tiene conocimiento del camino ni los motivos por los que se llegó a una decisión. Por ende, de no tomarse cautela necesaria al interpretar los resultados de una máquina de predicción, se puede incurrir

en decisiones erradas, lo que a la larga puede repercutir de manera perjudicial en el contexto donde estos sistemas se ejecuten. Si se considera, por ejemplo, un sistema de detección de cáncer a partir de una muestra de tejido, una idea prudente sería introducir un mecanismo que permita dilucidar incertidumbre asociada a la predicción, de manera de no validar de inmediato un resultado negativo, sino que, en caso de que exista una alta incertidumbre, solicitar una inspección y validación del resultado con un experto. De esta manera, se puede esclarecer una parte del funcionamiento de los métodos de DL, a partir de lo cual se pueden tomar decisiones mejor informadas, y evitar que se traten estos sistemas como caja negra.

Así mismo, lo que se propone como parte de esta memoria, es generar un modelo de predicción que pueda aportar con información adicional sobre los resultados que entregue, en la forma de una cuantificación de incertidumbre a través de un ensamble de modelos de redes neuronales o *deep ensemble* (DE). Esta información podrá apoyar de manera efectiva a los expertos, ya no sólo con la predicción misma, sino que con una estimación de la incertidumbre, lo que permitirá saber cuándo hace falta una revisión antes de validar un resultado.

1.5. Objetivos

De acuerdo los temas y problemáticas expuestos anteriormente, se plantean los siguientes objetivos para el desarrollo de esta memoria:

Objetivo General

Construir un modelo de predicción de parámetros estelares y estimación de incertidumbre de las predicciones usando *Deep Ensemble*.

Objetivos específicos

- Preprocesar la base de datos de líneas espectrales sintéticas para generar la entrada del modelo de predicción
- Diseñar, implementar y validar un modelo basado en redes neuronales, incluyendo el enfoque *Deep Ensemble*, para predecir los parámetros estelares.
- Definir e implementar métricas para medir la incertidumbre de las predicciones del modelo.
- Validar el modelo en un conjunto de líneas espectrales observadas y comparar con datos teóricos.

1.6. Contribuciones

Este trabajo analiza la efectividad de emplear una metodología de redes neuronales con un enfoque *deep ensemble* para la predicción de parámetros estelares a partir del espectro de una estrella. Con el objetivo de esclarecer la incertidumbre en los resultados arrojados, es que se entrega una estimación de la misma en base al error generado por una abundancia de predictores. De esta manera los astrónomos podrán contar con una herramienta adicional de análisis que les permitirá tomar decisiones mejor informadas y certeras.

Los resultados de este trabajo pueden ser replicados en el mismo ámbito de la astronomía para estudio de estrellas de clasificaciones y espectros diferentes. Y también, pueden usarse como apoyo en otras investigaciones en el ámbito de la inteligencia artificial, con problemáticas donde la cuantificación de la incertidumbre en la predicción pueda ser relevante, o simplemente necesaria.

El desarrollo completo del trabajo junto con la documentación relevante se pueden encontrar en el siguiente repositorio de GitHub: <https://github.com/chum1ngo/StellarParametraPredictionDeepEnsemble>.

CAPÍTULO 2

MARCO CONCEPTUAL

El aprendizaje automático (ML) es un área dentro de la inteligencia artificial que ha tenido un desarrollo excepcional durante las últimas décadas [Nath *et al.*, 2018]. Muchos de los destacados éxitos que se le atribuyen a la inteligencia artificial se caracterizan por emplear mecanismos de aprendizaje considerados dentro de este dominio. La relevancia de estos mecanismos es que, a diferencia de la algoritmia [Kaya, 2018], metodología antigua de computación, el aprendizaje automático no requiere una descripción de un procedimiento para llevar a cabo una tarea, lo único que necesita es una representación de los datos, y una descripción de la solución o respuesta correcta, con lo cual la máquina podrá establecer su propia forma de llegar a una solución, dotándola de una poderosa capacidad de interpretación.

Por otro lado, dentro del ML, existe un área particularmente prominente y que ha estado en boga durante el último tiempo debido a la versatilidad y efectividad que han demostrado los métodos que pertenecen a ella en distintos campos, como lo son la clasificación de imágenes, robótica, análisis y procesamiento de texto, sonido y vídeo, entre otros. Ésta es el área conocida como *deep learning* (DL) [LeCun *et al.*, 2015], cuya importancia radica en que los mecanismos de aprendizaje ya no necesitan una representación específica de la información para realizar la tarea. Ahora, en cambio, los datos se le pueden entregar a la máquina con poco o nulo procesamiento, además de que no se requiere una selección de los atributos relevantes para la tarea, la máquina puede aprender su propia representación de los datos y, más importante aún, generar su propio conocimiento, sin requerir inferencia del programador en el proceso. De esta manera, la máquina ya no solo tiene una fuerte capacidad de interpretación, sino que también de representación.

Esto significa un paso gigante en la automatización de tareas complejas, ya que generar una representación adecuada requiere de una selección meticulosa de las características que le permitan a la máquina desarrollar una tarea en particular. Este proceso usualmente conlleva investigación y toma de muestra de datos si es que los datos no existen previamente. Lo que puede ser especialmente engorroso si se requiere el apoyo de uno o varios expertos en el dominio, quienes con frecuencia son profesionales y cuentan con disponibilidad de tiempo limitada, además de representar un costo adicional para los proyectos.

En la presente sección se revisarán los fundamentos teóricos sobre los cuales se basa este trabajo. Empezando por los principios estadísticos que rigen la optimización de redes neuronales, seguido de las bases estructurales de las misma, y termina con introducción a la incertidumbre asociada a la predicción y una explicación del proceso de ensamblado de modelos.

2.1. Fundamentos

En función de establecer las bases que permitirán el desarrollo de la investigación e interpretación de los resultados, se revisarán aquellos conceptos pertinentes a la optimización de parámetros en las NN.

Suponiendo que existe un mecanismo subyacente de generación de los datos, cualquiera sea el contexto, el rol principal de la estadística en este caso consiste en aproximar dicho mecanismo, a través de la selección y ajuste de una familia modelos de predicción que pueda describir de manera generalizada la generación de dichos datos. Escogida una familia de modelos, la instancia que mejor aproxima al mecanismo original se obtiene a través del ajuste de parámetros, proceso también llamado entrenamiento, y que distingue al modelo óptimo del resto de su familia. A este modelo entrenado se le conoce como predictor. Por otro lado, el ajuste de parámetros puede transformarse en una tarea no trivial de acuerdo a la complejidad del problema, es por ello que se necesitan aproximaciones y técnicas que permitan abordar cada caso.

2.1.1. Minimización del Riesgo Empírico

Formalizando la idea anterior, de acuerdo a [Deisenroth *et al.*, 2020], suponiendo que tenemos n muestras $\mathbf{x}_i \in \mathbb{R}^D$, y las correspondientes etiquetas escalares $y_i \in \mathbb{R}$. Considerando el contexto del aprendizaje supervisado, tenemos los pares $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$. Dados estos datos, esperamos estimar un predictor $f(\cdot, \theta) : \mathbb{R}^D \rightarrow \mathbb{R}$ parametrizado por θ . De esta manera, nuestro objetivo es encontrar los parámetros θ^* tal que podamos ajustar los datos de lo suficientemente bien:

$$f(\mathbf{x}_i, \theta^*) \approx y_i, \forall i = 1, \dots, n \quad (1)$$

En adelante, consideraremos la notación $\hat{y}_i = f(\mathbf{x}_i, \theta^*)$ para describir la salida del predictor.

Considerando una etiqueta y_i de una instancia particular, además de la predicción correspondiente $\hat{y}_i$, necesitamos especificar una función de pérdida descrita como $\ell(y_n, \hat{y}_n)$ que tome el valor de la etiqueta y la predicción, y produzca un valor no negativo (*loss*) que represente qué tan grande es el error en una predicción particular durante la fase de entrenamiento. El objetivo para encontrar un buen vector de parámetros θ^* es minimizar la pérdida promedio en la muestra de n elementos.

Típicamente en ML suponemos que los datos con los que trabajamos son independientes e idénticamente distribuidos. La independencia entre los datos significa que un dato $(\mathbf{x}_i, y_i)$ no depende estadísticamente de otro dato $(\mathbf{x}_j, y_j)$, lo cual es importante, dado que implica que el promedio empírico es un buen estimador de la pérdida en los datos de entrenamiento.

Para el conjunto de datos de entrenamiento $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, introducimos la notación de los valores de entrada como la matriz $\mathbf{X} := [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top \in \mathbb{R}^{n \times D}$, y para las etiquetas correspondientes como el vector $\mathbf{y} := [y_1, \dots, y_n]^\top \in \mathbb{R}^n$. Con esta notación, la pérdida promedio se describe como:

$$\mathbf{R}_{\text{emp}}(f, \mathbf{X}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n \ell(y_i, \hat{y}_i), \quad (2)$$

donde $\hat{y}_i = f(\mathbf{x}_i, \boldsymbol{\theta})$. La Ecuación 2 se denomina el *riesgo empírico* y depende de 3 factores, el predictor f y los datos $\mathbf{X}, \mathbf{y}$. La estrategia general para el aprendizaje automático se llama *minimización del riesgo empírico*.

2.1.2. Función de Pérdida

Para el caso de regresión, se le conoce como función de pérdida al mecanismo que cuantifica el error que existe luego de efectuar una predicción sobre una instancia de entrenamiento, y al mismo tiempo, dicta la medida de ajuste sobre el predictor en base a esta cuantificación. Análogamente para el caso de clasificación, la función de pérdida penaliza de acuerdo a si una predicción es correcta o errada en la clase objetivo. Existen diversos métodos para aplicar una función de pérdida a una red neuronal tanto para la tarea de clasificación como de regresión [Goodfellow *et al.*, 2017], y elección de esta puede depender del tipo de penalización que se desee aplicar y de la distribución de los datos que se esté empleando. Debido a que en este caso se quiere determinar los parámetros de una estrella, propósito que coincide con la tarea de regresión, usaremos funciones de pérdida que nos permitan evaluar la calidad de una solución en términos de qué tan acertado o cercano es el valor de predicción respecto del valor real. A continuación se revisarán las funciones de pérdida que resultarán pertinentes a este estudio además de analizar sus ventajas y desventajas.

- **Mean Squared Error:** El error cuadrático medio (*Mean Squared Error* en inglés), de ahora en adelante MSE, consiste en tomar el cuadrado del error entre el valor real y la predicción, vale decir, el cuadrado de la diferencia entre ambos valores, y promediar este valor respecto de todas las instancias.

$$MSE = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

Una de las ventajas de esta función de pérdida consiste en su simplicidad, ya que solo tomamos como referencia la distancia entre los valores para tomar una decisión. Por otra parte, esta función supone que los datos de salida siguen una distribución normal, y dado que, se toma el cuadrado del error de la predicción como valor, esta función exacerbará y penalizará en mayor medida errores extremos respecto de aquellos que

se encuentren cerca de la media de la población, por lo que, se dice que es sensible a estos valores extremos, y en caso de que la distribución sea distinta de la normal, esta función de pérdida puede ser inadecuada.

- **Mean Absolute Error:** El error absoluto medio (*Mean Absolute Error*, en inglés), de ahora en adelante MAE, sigue una idea similar a la medida anterior, donde tomamos como referencia el promedio del valor absoluto del error de cada predicción.

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

La ventaja, en este caso, consiste en que esta función, dado que no toma por referencia al cuadrado del error, penalizará de manera proporcional cualquier error de predicción, por lo que no es sensible a valores extremos. Además de lo anterior, su simplicidad facilita la interpretabilidad del error.

2.1.3. Métricas de Desempeño

En una tarea de regresión, el objetivo es aproximar un valor, es decir, realizar una predicción. Para evaluar esta predicción, se debe comparar con el valor real, proceso que fué discutido en la sección anterior. Por lo que, para evaluar el desempeño del modelo que arroja estas predicciones, se distinguirá entre medidas utilizadas como función de pérdida y como métrica de evaluación del modelo, siguiendo esta estructura:

- MSE: Función de pérdida.
- MAE: Métrica de evaluación.

Por otra parte, la disparidad de escala entre las unidades de medida que nos interesan en este trabajo es significativa, por lo que se hace necesario tratar estos valores con un proceso de normalización antes de entrenar algún modelo con ellos. De acuerdo a ello, y en adelante, las métricas serán expresadas de acuerdo a la siguiente transformación de unidades:

- $T_{\text{eff}}^* = \frac{T_{\text{eff}}}{T_{\text{eff max}}}$, con $T_{\text{eff max}}$ el máximo valor de T_{eff} entre todos los datos del conjunto de entrenamiento.
- $\log g^* = \frac{\log g}{\log g_{\text{max}}}$, con $\log g_{\text{max}}$ el máximo valor de $\log g$ entre todos los datos del conjunto de entrenamiento.
- $v \sin i^* = \frac{v \sin i}{v \sin i_{\text{max}}}$, con $v \sin i_{\text{max}}$ el máximo valor de $v \sin i$ entre todos los datos del conjunto de entrenamiento.

2.2. Red Neuronal Artificial

Una red neuronal artificial (ANN, por sus siglas en inglés) es un tipo específico de modelo computacional, inspirada en estructuras naturales del mismo nombre que dan soporte al funcionamiento del cerebro. Así mismo, la unidad mínima de computación de la ANN es la neurona. Aprovechando la definición de [McCulloch y Pitts, 1943], desarrollada con profundidad en [Buduma *et al.*, 2022], la neurona toma por entrada o *input* una serie de datos $x_1, x_2, \dots, x_n$, cada uno de los cuales es multiplicado por un peso específico $w_1, w_2, \dots, w_n$ para después sumar los productos de dichas operaciones, lo que produce el denominado *logit* de una neurona, $z = \sum_{i=1}^n x_i w_i$. Regularmente, el *logit* también incluye el sesgo o *bias* b de la neurona, que es una constante de inicialización aleatoria. Finalmente, el *logit* es pasado por una función de activación f , la cual se definirá más adelante, que determina la salida ecuación $y = f(z)$, con lo cual, el valor y constituye el valor de salida u *output* de la neurona.

La formulación vectorial de la idea anterior se establece según lo siguiente: primero, los datos de entrada quedarán formulados en un vector $\mathbf{x} = [x_1, x_2, \dots, x_n]$, y los pesos de una neurona en relación a las neuronas de la capa anterior, en la forma $\mathbf{w} = [w_1, w_2, \dots, w_n]$, formas ante las cuales el producto punto equivale al *logit*. Adicionando el *bias* de la neurona, y considerando también la misma función de activación f , el *output* de la neurona queda formulado como:

$$y = f(\mathbf{w} \cdot \mathbf{x} + b) \quad (5)$$

2.2.1. Red Neuronal Feed Forward

La definición anterior entrega libertad para formular redes neuronales con diferentes estructuras, donde la salida de una neurona puede ser perfectamente un *input* para otra neurona. La estructura que nos interesa en este caso corresponde a la red neuronal *feed forward*, de la cual se puede ver un ejemplo en la Figura 3. En la figura, cada columna de nodos conforma una capa: la capa de entrada, la capa oculta, y la capa de salida. La capa de entrada es por la cual ingresa el *input* a la red neuronal, la capa oculta es la que computa ese *input* mediante la operación de la Ecuación 5. Esta capa lleva ese nombre debido a que, en la práctica, no es posible observar cómo opera los datos, sólo podemos conocer lo que entra y lo que sale de la red. Consecuentemente, cada una de las neuronas de este tipo de cada se denomina “unidad oculta”.

En la Figura 3 se puede ver un ejemplo de una red neuronal *feed forward* con una única capa oculta, aunque es necesario mencionar que es posible tener redes con más capas ocultas y no se requiere que todas tengan la misma cantidad de neuronas. La última capa es la capa de salida y los datos que arroja se pueden corresponder con la respuesta obtenida del problema, o datos entrada para otra estructura de computación dependiendo de la aplicación. Además, en la misma figura es posible notar que; la conexión entre una capa y la siguiente es densa, vale decir, todas las neuronas de una capa están conectadas con todas las de la siguiente; el

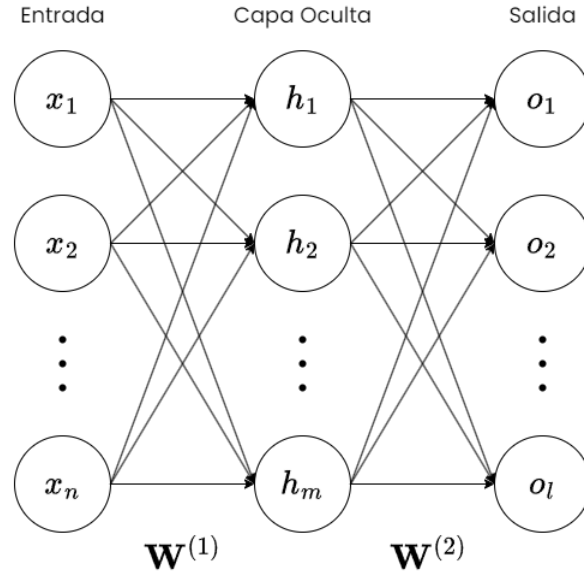


Figura 3: Red Neuronal Artificial *feed forward*. En esta representación, cada columna de nodos representa una capa de la red. La columna de más a la izquierda representa la capa de entrada de la red, la capa intermedia es la capa oculta, y la de más a la derecha representa la capa de salida. Los pesos aplicados a los datos al entrar en cada capa están representados por las matrices $\mathbf{W}^{(1)}$ y $\mathbf{W}^{(2)}$, cuyo superíndice indica la capa a los cuales cada matriz de peso está asociada. Fuente: Elaboración propia.

flujo de información va en una sola dirección (de izquierda a derecha), característica que le otorga el nombre a este tipo de estructura como “de avance” o *feed forward*.

Para describir la salida en este tipo de red $\mathbf{y} = [y_1, y_2, \dots, y_m]$ de cualquier capa con m neuronas se consideran: los n datos de entrada $\mathbf{x} = [x_1, x_2, \dots, x_n]$, los pesos descritos en forma matricial $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m]$ con $\mathbf{w}_i = [w_1, w_2, \dots, w_n]$ el vector columna de pesos de la neurona i siendo $i = 1, 2, \dots, m$, y $\mathbf{b} = [b_1, b_2, \dots, b_m]$ el vector que contiene el sesgo de cada neurona de la capa. La relación que describe el resultado de salida de la capa es:

$$\mathbf{y} = f(\mathbf{W}^T \mathbf{x} + \mathbf{b}) \quad (6)$$

Donde f es una función de activación aplicada a cada elemento del vector resultado. De acuerdo a la Ecuación 6, es fácil notar en la Figura 3, que los elementos $\mathbf{W}^{(1)}$ y $\mathbf{W}^{(2)}$ corresponden a los pesos de las neuronas de una capa descritos en forma matricial y que el superíndice de dichos elementos indica la capa a la cual pertenece la matriz de pesos. También es posible notar que la Ecuación 6 es muy similar a la Ecuación 5, sin embargo, es importante hacer la diferencia entre naturaleza matricial y multi-salida de la Ecuación 6 con respecto a la Ecuación 5.

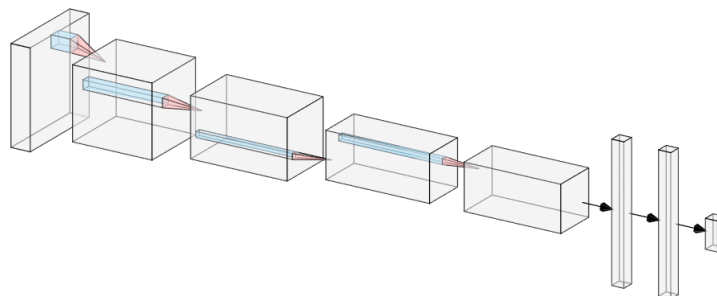


Figura 4: Representación de una Red Neuronal Convolutiva típica. De izquierda a derecha se tiene: el primer bloque representa la capa de entrada de los datos; los siguientes cuatro bloques representan capas convolucionales; los siguientes dos bloques corresponden a capas densas, donde se requiere una capa *flatten* para transformar los datos multidimensionales en vector; el último bloque es la capa de salida. Las conexiones entre las capas convolucionales están en color, y representan la convolución a través del *kernel* o filtro. Estas conexiones son pequeñas en comparación al bloque debido a que el *kernel* solo abarca una porción de los datos disponibles. La conexiones de las últimas tres capas son densas.

2.2.2. Red Neuronal Convolutiva

La red neuronal convolutiva (CNN, por sus siglas en inglés), es un tipo específico de ANN, y se caracteriza por incluir capas convolucionales, capaces de procesar datos en más de una dimensión. De acuerdo a [Goodfellow *et al.*, 2017], las capas convolucionales se estructuran en filtros o canales que procesan los datos de una capa anterior. Tradicionalmente, cada filtro de una capa tiene la misma arquitectura, pero captura aspectos diferentes de un mismo *input*. Para lograr esto, cada filtro tiene un *kernel* diferente asociado que transforma los datos. Este *kernel* actúa de manera similar a los pesos definidos en la red *feed forward*, pesando los datos de entrada y generando un dato de salida. Por otro lado, hay una diferencia fundamental respecto de la red *feed forward*, y es que en ella la conexión entre neuronas de una capa y la siguiente es densa, sin embargo, para el caso convolutivo cada neurona considera un conjunto acotado de todos los datos de entrada. Esto es lo que se conoce como el campo visual, y constituye el *input* propio de cada neurona. En consecuencia, el *kernel* debe de tener las mismas dimensiones que el campo visual de la neurona. En la Figura 4 se puede apreciar un ejemplo de red neuronal convolutiva.

El resultado de la transformación del *input* con el *kernel* es lo que se conoce como mapa de características (*feature map* en inglés). Esta denominación, en ocasiones, se utiliza como sinónimo de neurona para el caso convolutivo. Ahora bien, de lo discutido se sabe que el *kernel* es de menor tamaño que el dato de entrada, lo que presenta el problema de lograr integrar todos los datos en la siguiente capa. Esto se resuelve estableciendo un paso o *stride* de desplazamiento y agregando mas características al *feature map*. En la práctica, la metodología consiste en “barrer” todo el *input* utilizando el *kernel*, desplazándolo un paso a la vez y añadiendo una neurona a la salida de la capa con el resultado de cada operación. Formalizando la operación de convolución $w * x$ entre el *kernel* w y el *input* x para el caso unidimensional,

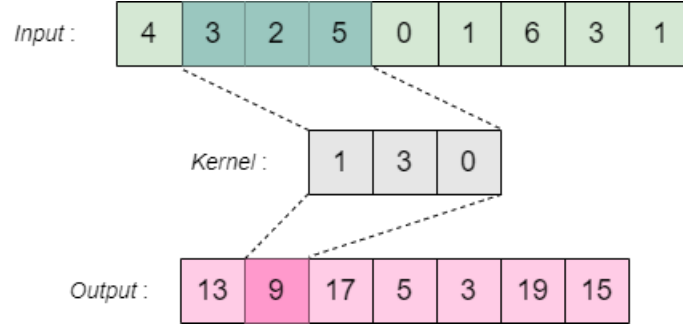


Figura 5: Ejemplo de Campo Visual y Kernel. La lista superior, en verde, representa el *input* y, en particular, la sección oscurecida es el campo visual; La lista en el medio representa el *kernel*; y la lista inferior representa los datos de salida (*output*) o mapa de características.

se tiene:

$$\begin{aligned} w * x[i] &= x * w[i] = \sum_{m=-\infty}^{\infty} x[m]w[i - m] \\ &= \sum_{m+r=i}^{\infty} x[m]w[r] \end{aligned} \quad (7)$$

La ecuación anterior plantea que el primer elemento del *input* x se relaciona con el último del *kernel* en la convolución. El índice que recorre x aumenta mientras que el índice que recorre w decrece y viceversa. Si bien se puede notar que para la Ecuación 7 se recorren todos los índices hasta el infinito, para acotar el *kernel*, basta con asignarle valor 0 a los pesos del *kernel* que se escapan de la dimensionalidad establecida. Lo mismo sucede para x . En la Figura 5 se puede encontrar una representación de la función de convolución para el caso unidimensional.

Con la descripción anterior, se puede establecer el *output* de una neurona mediante:

$$y = f(w * x - b) \quad (8)$$

Donde f corresponde a una función de activación de la neurona. En este caso se usará la función rectificadora lineal o *relu*, por sus siglas en inglés:

$$f(x) = \max(0, x) \quad (9)$$

donde $\max$ corresponde a la función que toma el máximo entre el valor de entrada x y 0. Esto se traduce, para la máquina, en el traspaso o no de información de una capa a otra dependiendo si existe la característica asociada o no.

La unidad oculta y la unidad de convolución son el corazón de la CNN y conllevan todo el procesamiento y proceso de aproximación al mecanismo generador de datos original, introducido

en la Sección 2.1. Estas unidades se estructuran en capas (*layers*, en inglés) que sistematizan y dividen el procesamiento de los datos en etapas.

Además de las capas convolucionales y ocultas o densas, existen otros tipos de capa relevantes a la CNN que complementan el trabajo de procesamiento de las capas ya mencionadas, a continuación, se listan y explican brevemente:

- **Capa de Entrada:** Tiene por única función adaptar la entrada de los datos a la red neuronal. Por lo mismo debe ser de la misma dimensión que los datos y no es entrenable.
- **Capa *MaxPooling*:** Al igual que la capa convolucional, esta capa tiene un *kernel* asociado, que permite sintetizar la información de una capa convolucional y reducir la dimensionalidad del mapa de características. Para esto selecciona el mayor valor dentro del campo de visión del *kernel* y lo entrega a la capa siguiente. Existen otros tipos de operación de *Pooling*, pero *MaxPool* es la más frecuentemente utilizada.
- **Capa *Dropout*:** Esta capa tiene por única funcionalidad el entregar regularización y permitir una mayor generalización de la red. Sea $\alpha \in [0, 1]$, este valor está presente para cada neurona de la capa y representa la probabilidad de que cada neurona se desactive con probabilidad α durante un ciclo de entrenamiento. Esto ayudará a la generalización de la red al forzar a la capa siguiente a crear generalizaciones que dependan de características diferentes a la de la neurona desactivada.
- **Capa de Regresión:** Es idéntica a la capa oculta en términos de unidades de procesamiento, sin embargo, lleva por función de activación la función *lineal*(x) = x . Se emplea al final de la CNN dado que, puede arrojar cualquier valor en el $\mathbb{R}$.

Cabe mencionar, que si bien las capas mencionadas son un estándar dentro del DL, la librerías principal a utilizar en este trabajo corresponde a *Autokeras* [Jin *et al.*, 2019], que implementa versiones personalizadas de algunas de las capas, ante lo cual se motiva al lector a familiarizarse en la respectiva documentación en <https://autokeras.com> para saber más.

Ahora bien, para que la red neuronal pueda entrenar y aprender de los datos, es necesario que ésta tenga un objetivo establecido en términos numéricos. Esto se traduce al ajuste u optimización de los parámetros luego de efectuar una predicción sobre un dato de entrenamiento. Para tener una noción de la medida de optimización a efectuar es necesaria la comparación entre una predicción y el valor esperado correspondiente. Esta comparación que se efectúa mediante la denominada función de pérdida, que establece el factor de error de la predicción en cada iteración de entrenamiento. En este caso corresponde al error cuadrático medio (MSE) y está dentro de las funciones de pérdida típicamente empleadas en problemas de regresión. Su estructura se revisará más adelante.

Una vez que se tenga la noción qué tan cerca se encuentra la predicción respecto del valor real, es necesario modificar los parámetros de la red de manera de generar una mejor predicción

cada vez que se introduce un dato de entrenamiento. Uno de los mecanismos para lograr esto es el de descenso por gradiente, que modifica los pesos de los mapas de características en cada capa de la ANN de acuerdo al tamaño del aporte de cada uno. La función de propagación del cambio respectivo en cada uno de los pesos se describe de la siguiente manera:

$$\Theta^{(t+1)} = \Theta^{(t)} - \eta_t \nabla J(\Theta^{(t)}) \quad (10)$$

donde se considera $\Theta^{(t+1)}$ como el nuevo valor de un peso de la matriz de características, como la suma entre el valor del peso en la iteración anterior y el negativo de la dirección del gradiente multiplicado por η_t , parámetro de la red conocido como tasa de aprendizaje (*learning rate*, en inglés).

Cabe destacar que la función de optimización de los pesos de la ANN mencionada no necesariamente coincide con la función de optimización empleada en las CNN de este trabajo, pero constituye el punto de inicio para la comprensión de las mismas, he ahí su relevancia en esta sección. Descripciones de algunas funciones de optimización empleadas en este trabajo se pueden encontrar en [Kingma y Ba, 2014] y [Loshchilov y Hutter, 2017].

2.3. Cuantificación de la incertidumbre

Durante el proceso de resolver un problema de ingeniería, el ingeniero o la ingeniera debe tomar decisiones sobre las variables relevantes en el contexto del problema. Estas variables deben considerarse en el modelo que las relaciona (de una manera matemática e idealizada) con un fenómeno físico, que suele ser difícil de medir contextos de producción, y sobre el cual se desea realizar una predicción o determinar su comportamiento. Son relevantes las suposiciones acerca de la incertidumbre asociada a dichas variables, ya que determinan la confiabilidad de la información que entrega un modelo sobre un proceso.

Por un lado, existe un tipo de incertidumbre intrínseca con respecto a los datos que se van a usar para alimentar al modelo. Esta incertidumbre viene dada por la precisión de la medición, y puede estar influenciada por el instrumento o la metodología con la cual se realiza, determinando la magnitud del error.

Por otro lado, dentro del contexto del problema, puede existir un universo de técnicas que modelen el fenómeno físico, por lo que es relevante considerar la incertidumbre que existe con respecto al espacio posible de soluciones subyacentes a la técnica empleada. Lo que, en concreto, se evalúa como la brecha entre el mejor modelo posible dentro de un espacio, y el fenómeno físico que genera los datos.

Por último, ya efectuada la elección del modelo, también existe, como ya se habló en secciones previas, parámetros que determinan a un modelo dentro de la familia a la que pertenece, sobre los cuales también existe incertidumbre. Ésta se describe como la distancia entre la mejor solución encontrada y la mejor solución posible, dentro del espacio de soluciones que abarca el modelo.

Estas fuentes de incertidumbre se pueden clasificar de acuerdo a si se pueden abordar dentro del problema a resolver o no [Der Kiureghian y Ditlevsen, 2009], por lo cual se identificarán las fuentes de incertidumbre que incumben al problema en cuestión en el presente trabajo junto con la naturaleza de su clasificación:

- **Incetidumbre Aleatoria:** Corresponde a aquellas fuentes de incertidumbre que conllevan una connotación aleatoria o estocástica en el proceso subyacente que se desea modelar, y que no se pueden abordar debido a su naturaleza impredecible [Dodwell, 2023]. Para el presente trabajo, se considerarán como tales:
 - **Error de medición:** Corresponde al error de precisión inminente del instrumento de medición (para el caso, un espectrógrafo) al tomar la muestra del espectro de cada estrella. Además de lo anterior existe un factor aleatorio en el comportamiento mismo de la estrella, con lo cual las mediciones tomadas pueden variar considerablemente de un período a otro.
- **Incetidumbre epistémica:** Corresponde a aquellas fuentes de incertidumbre que provienen de la falta de conocimiento o datos acerca del proceso que se quiere modelar. Son abordables dentro del proceso de modelado y análisis, lo que puede ayudar a ajustar, acotar y/o explicar los resultados del modelo empleado [Dodwell, 2023]. Para el presente trabajo se tendrán en cuenta las siguientes:
 - **Error de aproximación:** Mencionado anteriormente, consiste en la distancia entre el proceso físico que se quiere modelar, y la mejor aproximación de este que se pueda lograr en un espacio de funciones determinado. La metodología para reducir este error consiste en buscar la mejor arquitectura posible para la predicción de parámetros estelares con *Autokeras*. Librería de aprendizaje automático que facilita la implementación de redes neuronales mediante generación automática de modelos.
 - **Error de estimación:** Como se mencionó anteriormente, existe una diferencia en cuanto al mejor modelo posible dentro de una familia de éstos y el mejor que se logra encontrar de acuerdo a los datos obtenidos del proceso original. La metodología que permitirá reducir este error, y que constituye uno de los temas principales de este trabajo, consiste en la generación de ensambles de modelos, que permite no solo reducir el error de estimación, si no que permite estimar la incertidumbre de una predicción.

2.4. Descomposición sesgo-varianza

Una formulación matemática de la idea anterior corresponde a la descomposición en sesgo y varianza del error de una predicción. Suponiendo una regresión lineal, donde la definición del

error corresponde al error cuadrático, considerando el par de datos observados (x, y) generado a partir función de probabilidad f , que representa al mecanismo subyacente generador de datos, y que la relación entre las variables corresponde a $y = f(x) + \epsilon$, donde ϵ es un factor de error que sigue una distribución normal $\epsilon \sim N(0, \sigma_\epsilon)$. Además, considerando f_S un predictor entrenado sobre una muestra de datos S . La formulación del error para el predictor f_S según [Neal *et al.*, 2018] es:

$$\text{Error}(x) = \mathbb{E}[(y - f_S(x))^2] \quad (11)$$

Lo cual se puede descomponer en los siguientes términos:

$$\text{Error}(x) = \underbrace{\mathbb{E}[(f_S(x) - \mathbb{E}[f_S(x)])^2]}_{\text{varianza}} + \underbrace{(\mathbb{E}[f_S(x) - f(x)])^2}_{\text{sesgo}^2} + \underbrace{\sigma_\epsilon^2}_{\text{ruido}} \quad (12)$$

En la fórmula anterior podemos distinguir tres elementos relevantes:

- El sesgo, describe la diferencia entre la predicción y el valor real.
- La varianza, que describe la distancia entre la predicción y su valor esperado.
- El ruido, que constituye la fuente de error irreducible de la predicción

De estos elementos, aquel que resulta atractivo para este estudio es la varianza. Notar que el estimador entrenado f_S está condicionado por el conjunto S con el que fue entrenado, por tanto la componente de la varianza entrega información acerca de cuánto cambia el error de predicción conforme cambie el conjunto de entrenamiento S .

2.5. Intercambio entre sesgo y varianza

Una idea para comprender la sección anterior, corresponde a considerar todos los posibles predictores (dentro del espacio de soluciones) que un algoritmo de ML puede arrojar para generar una sola predicción. Suponiendo además, que se tuviesen infinitos conjuntos de datos para entrenar cada instancia del modelo, de manera que cada predictor sea independiente de los demás. La premisa es que, si el espacio es demasiado grande, entonces la varianza asociada sería grande, dada la libertad con la que se pueden generar predictores. Pero adicionalmente, el error de aproximación sería pequeño, dado que consideramos un espacio de soluciones que podría estar más cerca de el mecanismo generador de datos original.

Por el contrario, si se acota el área de búsqueda de los predictores, la varianza disminuirá, pero también aumentaría el error de aproximación, dado que se podría estar coartando al algoritmo de generar un predictor cercano al el mecanismo original.

Esto es lo que se conoce como intercambio entre sesgo y varianza que, en resumidas cuentas, nos dicta que al aumentar el espacio de búsqueda, hay más posibilidad de encontrar un mejor predictor, pero que a su vez, también es más fácil toparse con sobre-ajuste.

La perspectiva que puede aportar DE respecto de la cuestión anterior, reside en su aporte para reducir la varianza de la predicción. De acuerdo a [Arvind *et al.*, 2023], se comprueba que es posible reducir la varianza asociada a la predicción cuando la cantidad de predictores del ensamble tiende a infinito.

2.6. Ensamble de modelos

Típicamente antes de tomar una decisión en algún ámbito es deseable contar con la mayor cantidad de información posible y provenga de diversas fuentes para evitar sesgo. Análogamente, con los modelos de predicción, es posible combinar la información arrojada por distintos modelos para mejorar el desempeño general. De ahí proviene la idea de ensamblar modelos de predicción, de manera de combinar sus fortalezas para lograr un modelo más robusto en comparación a cada uno de los modelos por separado. Para predicciones de modelos respecto a la tarea de regresión existe más de una metodología, sin embargo, para este trabajo nos concentraremos en las siguientes, descritos en [Hastie *et al.*, 2009]:

- **Bagging:** Si se entrenasen varios modelos con la misma arquitectura sobre el mismo conjunto de datos, probablemente el resultado sería que los modelos predijesen lo mismo en instancia nunca antes vistas. Es por ello que esta técnica lo que intenta hacer es diferenciar los modelos entregándoles subconjuntos de entrenamiento diferentes a cada modelo obtenidos del conjunto de datos original.
- **Bootstrapping:** Para evitar la reducción del tamaño del conjunto de entrenamiento de cada modelo, y al mismo tiempo diferenciar cada conjunto (en la medida de lo posible), esta técnica toma muestras aleatorias con reemplazo desde el conjunto de datos original para generar conjuntos de entrenamiento con distribuciones diferentes.
- **Weighted Average:** Sean n modelos destinados para el ensamble, dada la predicción $\hat{y}_i$ de cada modelo i , con $i = 1, 2, \dots, n$ de cada modelo, se combinan mediante el promedio pesado (*weighted average*, en inglés) para cada instancia de la siguiente manera [Pawlikowski y Chorowska, 2020]:

$$\bar{y} = \sum_{i=1}^n \hat{y}_i \cdot p_i \quad (13)$$

El peso p_i asignado a cada modelo está relacionado la métrica de error MAE obtenida de la predicción sobre un conjunto de prueba apartado del conjunto de datos inicial con este solo propósito. Considerando los mismos n modelos para el ensamble, sea e_i el valor de la métrica

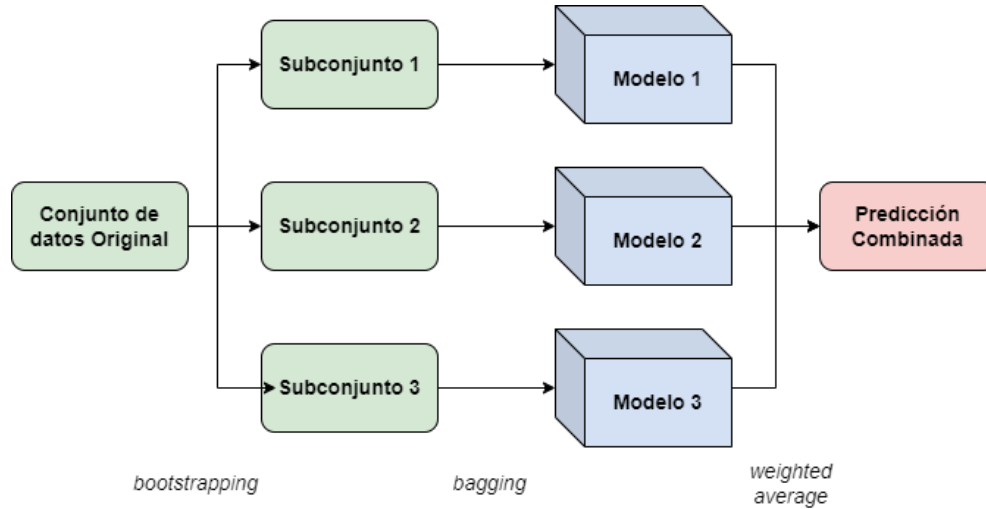


Figura 6: Representación del flujo de trabajo del ensemble de modelos.

desempeño de un modelo cualquiera, con $i = 1, 2, \dots, n$, valor del peso p_k de un modelo k está dado por:

$$p_k = \frac{1 - e_k}{\sum_{i=1}^n 1 - e_i} \quad (14)$$

Debido a que la métrica de desempeño indica una mejora cuando el error entre la predicción y el valor esperado se reduce y al escalamiento de los datos de salida, este valor tiende a ser lo más pequeño posible. La Ecuación 14 considera esta situación, de manera que el modelo más pesado es el que logre el valor más cercano a 1 en el numerador de la ecuación. Finalmente, el peso de cada predictor se divide con respecto a la suma de todos los pesos a modo de escalar el resultado.

Para describir el flujo relacionado a estas metodologías, en la Figura 6 se describe el orden en que son utilizadas. Primero, desde el conjunto de datos original se toman muestras de datos con *bootstrapping*. Después estos subconjuntos son asignados a diferentes instancias de un modelo de CNN, proceso conocido como *bagging*. Por último, y después del entrenamiento, los resultados de los distintos modelos entrenados son promediados mediante *weighted average*.

Adicionalmente, una de las ventajas de esta metodología es tener múltiples observaciones sobre una misma predicción, dadas por los modelos entrenados con conjuntos de datos, en principio, diferentes. Esto genera que los resultados puedan variar enormemente de un modelo entrenado a otro, lo que nos puede dar una idea de la incertidumbre que presentan los modelos ante ciertas instancias del entrenamiento, incertidumbre que puede estar dada tanto por la poca representatividad que tienen algunas instancias puedan tener, como por la variabilidad que puedan presentar respecto de una misma clase o, en este caso, respecto a un mismo valor

esperado o una vecindad de este. Del mismo modo, es posible capturar dicha incertidumbre determinando la diferencia entre las predicciones ponderadas s_i para una misma instancia. Para esto se utilizará la diferencia cuadrada de cada predicción ponderada:

$$s_i = \sqrt{\sum_{i=1}^n (\bar{y} - \hat{y}_i)^2 \cdot p_k} \quad (15)$$

2.7. Estado del Arte

El problema de la predicción de parámetros estelares a partir del espectro observado ha sido ampliamente abordado en la literatura. Existen varios métodos que se han utilizado y optimizado con anterioridad además de herramientas, así como también se han generado nuevos métodos recientemente, sobre todo bajo el alero del *deep learning*. A continuación se hará una revisión de algunas de las metodologías empleadas en las últimas décadas.

Aproximaciones que se centran en comparar y minimizar la distancia entre espectros sintéticos y observados se pueden encontrar en [Arcos *et al.*, 2018] y [Allende Prieto, 2004] que desarrollan sus estudios en líneas específicas de absorción. Incluso existen grandes proyectos de estimación masiva de parámetros en catálogos de estrellas que ocupan este tipo de aproximación, como en [Katz *et al.*, 1998], [Lee *et al.*, 2008], [Wu *et al.*, 2011] o [Xiang *et al.*, 2015]. Dichos trabajos se enmarcan dentro de los métodos de minimización de métrica (MDM) [Bailer-Jones, 2001], y no necesariamente contemplan el ajuste de un sistema de estimación.

Con la publicación de nuevos catálogos masivos de espectros ([Majewski *et al.*, 2017], [Ji *et al.*, 2023], [Royer *et al.*, 2024]), se hace necesario el uso de nuevas herramientas que aborden la alta resolución con la que se miden estos espectros, debido a que esta cualidad genera alta dimensionalidad y costo computacional. Desde la perspectiva de la alta dimensionalidad, técnicas que permiten la extracción de rasgos característicos del espectro resultan atractivas para esta tarea. Por ejemplo, la técnica de análisis de componente principal (PCA, por sus siglas en inglés) ha tenido un rol fundamental en el análisis de espectros. Esta técnica que permite reducir la dimensionalidad de los espectros capturando sólo los rasgos que entregan más información para luego procesarlos. En [Folkes *et al.*, 1996] se emplea esta técnica para pasar a un proceso de filtrado y comparación con espectros conocidos. También, [Singh *et al.*, 1998] usa la reducción de dimensionalidad para finalmente determinar los parámetros mediante una red neuronal de dos capas ocultas. Otra técnica empleada por [Li *et al.*, 2014] implica la extracción de características mediante el algoritmo LASSO [Tibshirani, 1996], para luego realizar una regresión mediante un algoritmo de soporte vectorial.

También, aproximaciones más recientes dentro del ML más tradicional basadas en ensambles de árboles de decisión, se pueden encontrar en [Xiang *et al.*, 2021] empleando *catboost* [Dorogush *et al.*, 2018] después de un proceso de selección de características, o también en

[Liang *et al.*, 2022] que emplea un mecanismo de PCA más *Light Gradient Boosting Machine* [Ke *et al.*, 2017]. Sin embargo, estas metodologías todavía requieren la reducción de dimensionalidad debido al alto coste computacional.

Por otro lado, existen técnicas que permiten abordar con mayor efectividad la abundancia de datos en los catálogos de estrellas, como es el caso de las redes neuronales. En [Fabbro *et al.*, 2018] se propuso una red neuronal llamada StarNet, arquitectura entrenada y probada tanto con espectros sintéticos como observados, que permite realizar predicciones de algunos parámetros empleando un rango amplio del espectro como dato de entrada. En otro trabajo de [Zhang *et al.*, 2019], se emplea la misma red StarNet, esta vez extendiendo la predicción incluyendo abundancias de ciertos elementos. También, en [Li y Lin, 2023] se propone la red StarGRU que utiliza mecanismos recurrentes y atencionales en la estimación de parámetros. Otra aproximación, en [Ness *et al.*, 2015] entrena una red neuronal generativa llamada The Cannon que genera etiquetas a partir de espectros. Otros trabajos incluyen [Ksoll *et al.*, 2020], donde se utiliza una red neuronal condicional invertible, o [Cai *et al.*, 2023], que utiliza DenseNet, una arquitectura de red neuronal proveniente del área de *computer vision*, pensada para la clasificación de imágenes 2D, pero que fué adaptada a la estimación de parámetros sobre espectros 1D. Incluso, existen herramientas disponibles basadas en redes neuronales como *AstroNN* [Leung y Bovy, 2019], que lidia con el problema de datos de entrenamiento incompletos y/o ruidosos, o *The Payne* [Ting *et al.*, 2019], que permite la estimación de parámetros a través de la interpolación de valores en grillas de valores de parámetros conocidos.

De los trabajos anteriores se puede rescatar el uso recurrente de espectros sintéticos en la comparación con espectros observados o en el entrenamiento de modelos de predicción. También, la alta dimensionalidad de los espectros resulta en un problema frecuente, lo que incentiva el uso de un sector acotado del espectro; mecanismos de reducción de dimensionalidad; o en su defecto, mecanismos que permitan una estimación rápida de parámetros, como es el caso de las redes neuronales.

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

Hoy en día existen diversos métodos para la estimación de los parámetros estelares de una estrella a partir del espectro observado. Muchos de ellos conllevan la comparación con métodos de ajuste entre un conjunto de espectros teóricos y el observado [Decin *et al.*, 2004], otros conllevan la especificación de parámetros atmosféricos para la generación sistemática de espectros sintéticos hasta lograr un nivel de similaridad que satisfaga cierta métrica de comparación [Shkedy *et al.*, 2007].

Sin embargo, estos métodos requieren un esfuerzo prohibitivo en cuanto al procesamiento que requieren para grandes volúmenes de datos, ya sea por que el método para comparar espectros es costoso, o lo es el método para generar espectros sintéticos. En este sentido es que nacen aproximaciones desde la perspectiva de la inteligencia artificial que permitan producir estimaciones rápidas a un mucho menor costo.

El presente trabajo concentra sus esfuerzos en comprobar la efectividad de la estimación de parámetros estelares mediante una metodología que integra técnicas del DL y ML tradicional, denominada *Deep Ensemble* (DE), que combina la predicción de varias CNN para generar una predicción única, y que además, entrega su respectiva varianza asociada, lo que permite tener una interpretación más robusta al tener una noción de incertidumbre en la predicción. Además de lo anterior, resulta relevante destacar una parte crucial del proceso de entrenamiento de las CNN, y consiste en que estas se entrenarán sobre un conjunto de datos sintéticos antes de predecir sobre datos observados.

La elección del modelo de CNN que se usará como pieza del ensamble de modelos, se realizará mediante la herramienta *Autokeras* [Jin *et al.*, 2019], que automatiza la selección de las característica de la red neuronal dentro de un espacio de búsqueda. El entrenamiento de cada CNN del ensamble se efectuará mediante una metodología de generación de muestras de datos diferentes, con el objetivo de que cada CNN no entrene sobre el mismo conjunto de datos y los parámetros de cada red no sean iguales al finalizar.

Posteriormente se realizará una estimación de parámetros estelares sobre un conjunto acotado de espectros de estrellas tipo Be ya estudiados en [Arcos *et al.*, 2018] y se compararán los resultados.

En los puntos siguientes se describirán los procesos que constituyen el flujo de trabajo. En la Sección 3.1, se describe el proceso de generación de los datos sintéticos. En la Secciones 3.2 y 3.3 se detallan las librerías a utilizar y una red neuronal canónica (no optimizada) propuesta sobre los datos iniciales respectivamente. En la Sección 3.4 se detalla el proceso de encontrar una red neuronal optimizada con *Autokeras*. En la sección 3.5 se describirá el proceso de ensamblado de las predicciones y estimación del error. En la Secciones 3.6 y 3.7 se explicará el proceso de estimaciones realizadas sobre datos reales, y se dará una introducción a la

evaluación de posibles errores encontrados en la predicción respectivamente.

3.1. Generación de espectros sintéticos

La generación de espectros sintéticos para el entrenamiento de la red neuronal corresponde al primer paso del flujo de trabajo para la determinación de los parámetros estelares. La generación de estos espectros tiene, de por sí, sus propios pasos de ajuste debido a que debe pasar por varios programas de software, los cuales son Tlusty [Hubeny, 1988], Synspec y Rotin [Hubeny y Lanz, 2017].

3.1.1. Tlusty

Tlusty [Hubeny, 1988] es un herramienta de software desarrollado con la finalidad de producir modelos de estrellas a partir de parámetros entregados al programa. Si bien, los modelos son esenciales para la generación de los respectivos espectros, en este trabajo no se generarán modelos en específico para la tarea, sino que, se utilizarán los modelos establecidos en el trabajo [Lanz y Hubeny, 2007] que recopila un conjunto de modelos sintéticos de estrellas de tipo O y B descritos por una grilla de parámetros.

Los autores del trabajo mencionado legaron modelos atmosféricos para las configuraciones mostradas en la Figura 7 de estrellas tipo espectral B. Esto, en conjuntos de configuraciones específicas según la metalicidad atribuida al conjunto de modelos de las estrellas con respecto nuestro Sol. De estos, se emplearán únicamente aquellos que tienen una relación de metalicidad igual a 1, es decir, los modelos de aquellas estrellas con metalicidad igual a la del Sol. A partir de estos modelos seleccionados, se generarán los espectros sintéticos.

Después de detallar todos los parámetros que se emplearán para la generación de espectros, la Tabla 1 resume la especificación del rango y paso de los modelos entregados en [Hubeny, 1988] que se utilizarán en este trabajo.

Más información acerca del funcionamiento del programa, guías y referencias, pueden ser encontradas en el sitio web de Tlusty, <http://tlusty.oca.eu>.

3.1.2. Synspec

Synspec [Hubeny y Lanz, 2017] es la herramienta que realiza la síntesis de los espectros de luz a partir de modelos atmosféricos. Permite diferentes configuraciones, entre ellas el rango en el cual se genera el espectro y la precisión del mismo. Tiene por característica el generar dos espectros. El primero de ellos representaría al espectro normal de la estrella modelada, y el segundo, un espectro continuo que servirá de referencia para normalizar el espectro anterior.

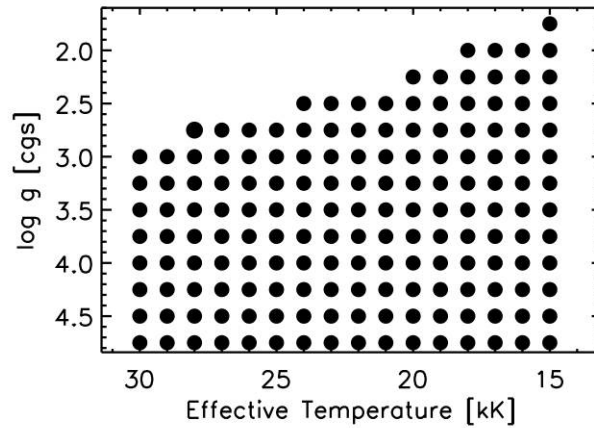


Figura 7: Grilla de Parámetros: Estrellas de tipo B. En la grilla se pueden comprobar las configuraciones de parámetros disponibles de modelos generados en [Lanz y Hubeny, 2007]. Cada punto representa un modelo con una T_{eff} y un $\log g$ específicos. Fuente: [Lanz y Hubeny, 2007].

En la Figura 8 se puede apreciar un ejemplo de ambos espectros en el rango de longitud de onda visible al ojo humano, y en la Figura 9 se magnifica el espectro sobre la región de interés a este estudio.

La información específica sobre la configuración del software previa a la generación de los espectros se encuentra en el Anexo G. Más información acerca del funcionamiento del programa, además de guías y referencias, pueden ser encontradas en el sitio web de Synspec, <http://tlusty.oca.eu/Synspec49/synspec.html>.

3.1.3. Rotin

La rotación de una estrella tiene un efecto importante sobre el espectro estelar observado. Por ello, es necesario simular las variaciones que puede tener el espectro de acuerdo a la cada rotación posible para la estrella.

Para lograr lo anterior, se utilizará la herramienta Rotin [Hubeny y Lanz, 2017], que toma como dato de entrada los dos espectros generados por Synspec y una velocidad de rotación de referencia (entre otros parámetros de configuración), para generar una simulación de cómo se vería el espectro convolucionado para dicha velocidad de rotación. Junto con el proceso de rotado también se genera una normalización del espectro, por tanto el espectro final ya no es similar al espectro real, sino que está normalizado con respecto al continuo, lo que permite visualizar con claridad las líneas de absorción existentes. Este proceso se puede repetir con múltiples velocidades de rotación para un mismo espectro, lo que puede generar multitud de líneas espectrales sintéticas. En la Figura 10 se pueden apreciar 3 convoluciones generadas con Rotin, para 3 valores de rotación diferentes de un mismo espectro de Synspec.

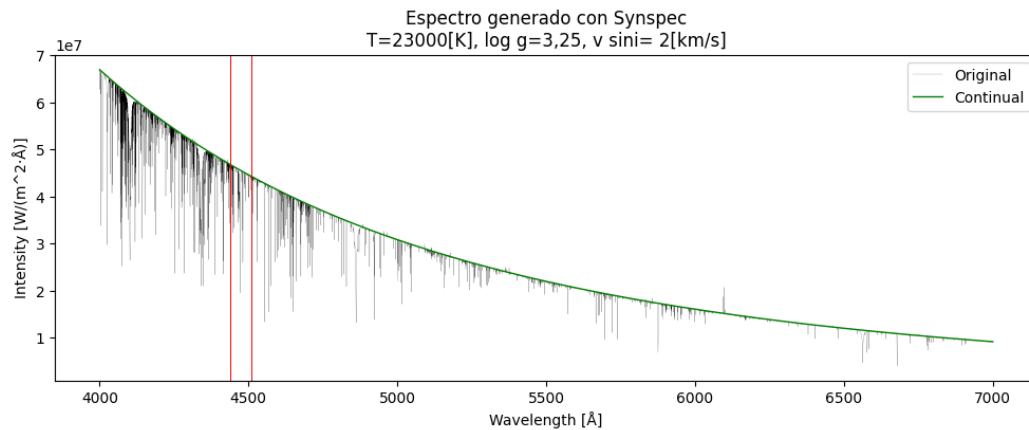


Figura 8: Espectro generado con Synspec. Ejemplo de espectro generado con Synspec en el rango de longitud de onda visible al ojo humano. Las líneas rojas se encuentran en los valores de $4440[\text{Å}]$ y $4510[\text{Å}]$, y denotan la región del espectro de interés a este trabajo. Fuente: Elaboración propia.

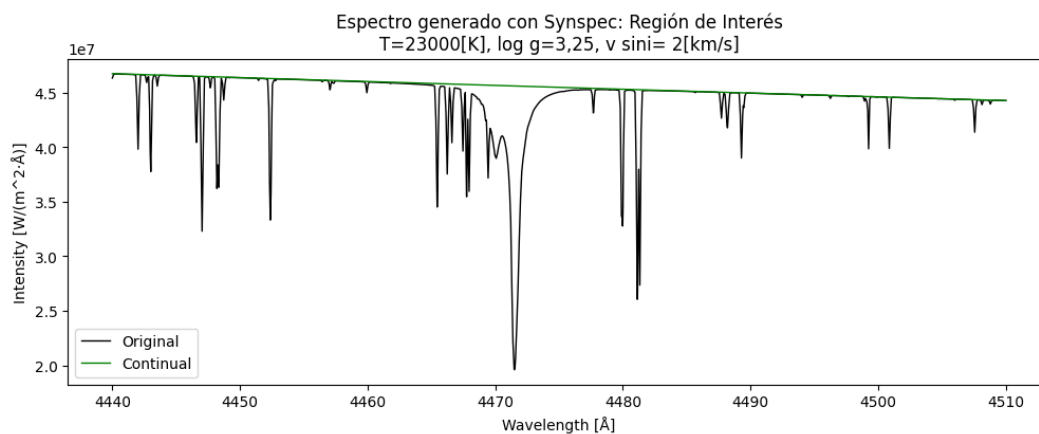


Figura 9: Espectro generado con Synspec, Región de Interés: Ampliación de la región de interés dentro del espectro visible. Al centro de la figura se puede apreciar con mayor claridad la línea de absorción a estudiar. Fuente: Elaboración propia.

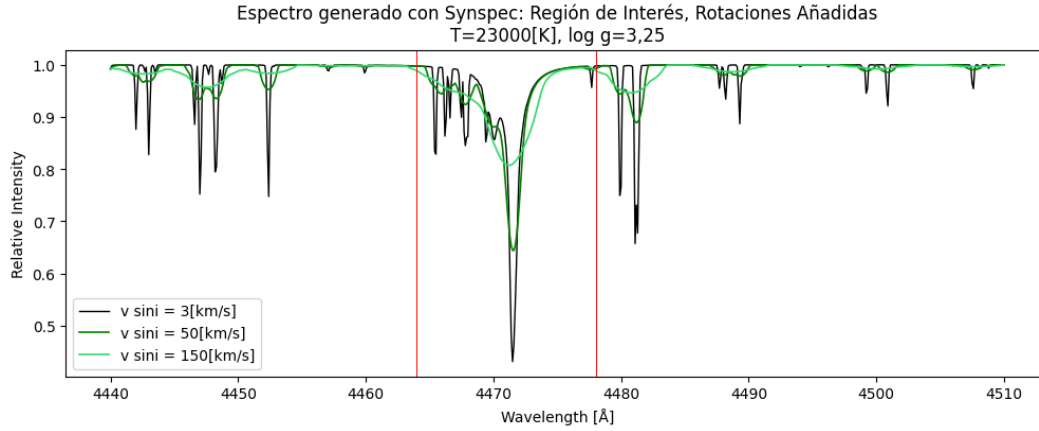


Figura 10: Espectro Generado con Rotin. Ejemplo de espectros generado con Rotin para diferentes $v \sin i$ a partir de los espectros continuo y normal entregados por Synspec. Se puede apreciar que, a diferencia de las figuras 8 y 9, la magnitud de la intensidad lumínica ahora está en torno a 1 y que ya no tiene unidad de medida dado que los datos están normalizados respecto al espectro continuo. La región que yace entre las líneas en rojo acota con mayor precisión la línea de absorción a estudiar.

El detalle de los parámetros de configuración del programa se encuentran en el Anexo H. Más información acerca del funcionamiento del programa, y referencias puede ser encontrada en el sitio web que comparte con Synspec, <http://tlusty.oca.eu/Synspec49/synspec.html>.

3.1.4. Línea de absorción del Helio

Para efectuar la predicción de los parámetros estelares, en el presente trabajo se estudiará una de las líneas de absorción del Helio ionizado I (HeI), elemento que caracteriza a las estrellas de tipo O y B, debido a que el ion sólo se encuentra en estrellas muy calientes. En específico, se analizará la línea con longitud de onda cercana a los 4471[Å], que se corresponde con la energía necesaria para efectuar la transición de un electrón del HeI desde un determinado nivel a uno inferior. El rango empleado para acotar los espectros de absorción sintéticos a esta línea va desde los 4464[Å] a los 4478[Å].

3.1.5. Configuración General para la Generación de Espectros

De acuerdo a la sección anterior, existen dos programas que configurar para la generación de espectros. En particular, considerando los parámetros de la Tabla 1, que resume los parámetros relevantes para Synspec y Rotin, con los que se podrán determinar la cantidad de espectros a generar.

Respecto a Synspec, el parámetro de configuración más relevante para el presente trabajo, consiste en el rango de longitud de onda en el cual se van a generar los espectros. Para este

Model	Thrusty/step
$T_{\text{eff}}[\text{K}]$	15000-30000/1000
$\log g[1]$	1.75-4.75/0.25
Solar metallicity	$Z/Z_0 = 1,0$
$v \sin i[\text{km/s}]$	3-450/3

Tabla 1: Resumen de parámetros, Synspec y Rotin.

trabajo, considerando la posición de la línea de absorción del HeI, y el hecho de que queremos perder la menor cantidad de información posible al acotar la línea de absorción, por lo que el rango de generación de los espectros considerará longitudes de onda desde 4440[Å] a los 4510[Å].

Para Rotin, el parámetro que resulta relevante mencionar corresponde a la velocidad de rotación $v \sin i$ asociada a la estrella. Para el presente trabajo se emplearán valores que van desde los 3[km/s] a los 450[km/s] con paso igual a 3[km/s], lo que da un total de 150 espectros generados por cada modelo atmosférico. Los valores empleados para el rango se explican desde el punto de vista físico, donde es posible tener estrellas que rotan a velocidad casi nula, cercana a cero, hasta velocidades muy altas que, sin embargo, tienen un límite dado por el punto en que la fuerza centrífuga, provocada por la rotación en el ecuador estelar, excede a la fuerza gravedad, lo que produce achatamiento, e incluso, expulsión de material desde la estrella. Esto explica un rango de velocidad inferior a dicho límite, donde la estrella tiene un comportamiento predecible.

Con esta información y la de los puntos anteriores, el flujo que permite la generación de los espectros sintéticos, resumido en la Figura 11 consiste en lo siguiente:

1. Obtener los modelos de estrellas tipo B con metalicidad similar a la de la Tierra.
2. Generar espectros con Synspec entre las longitudes de onda 4440[Å] y 4510[Å].
3. Generar el conjunto de convoluciones con Rotin tal que, la velocidad de rotación asociada a la convolución, está en el rango de 3[km/s] a 450[km/s] (inclusive) con una diferencia de 3[km/s] entre un valor y el siguiente.
4. Recortar los espectros generados entre las longitudes de onda 4464[Å] y 4478[Å].
5. Unificar los espectros generados con sus respectivos parámetros estelares para el entrenamiento de la red neuronal en un formato único.

A partir de la unificación de los datos se podrá distribuir fácilmente la separación de los datos en conjuntos de entrenamiento, validación, y prueba (*test*). Metodología muy empleada dentro del área de ML para validar la efectividad del modelo escogido y además tener una cuantificación del error en ámbito de producción [Hastie *et al.*, 2009]. En la Tabla 2 se puede ver la cantidad de datos para cada conjunto.

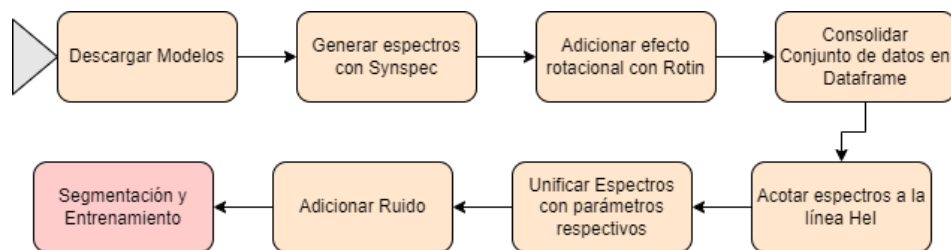


Figura 11: Flujo de Generación de espectros sintéticos. El flujo comienza por el preprocesamiento inicial de los datos y termina cuando estos están listos para ser usados para entrenar redes neuronales.

Train (60 %)	Validation (20 %)	Test (20 %)	Total (100 %)
14670	4890	4890	24450

Tabla 2: Separación del conjunto de datos.

Sobre el conjunto de datos generado, se podrá añadir ruido a los datos, para simular inestabilidad de la medición con el espectrógrafo. El ruido añadido para ciertas instancias de entrenamiento sigue una distribución $\sim N(0, 1)$ escalado a 0.01, lo que representa también el nivel de ruido esperado de los datos observados. En la Figura 12 se puede apreciar una representación de un espectro sintético, con y sin ruido añadido.

Cabe mencionar que la experimentación realizada arrojó diferencias importantes al utilizar instancias de espectro sintético con diferente rango de longitud de onda, ante lo cual, y para facilitar la comprensión de este documento, se define la siguiente nomenclatura para los datos:

- **De rango extendido (o no acotados):** entre los 4440[Å] y 4510[Å].
- **Acotados:** entre los 4464[Å] y 4478[Å].

3.2. Implementación de Redes Neuronales

Como ya se mencionó anteriormente, el objetivo de este trabajo es verificar el desempeño de un ensamble de modelos en la predicción de parámetros estelares a partir de espectros. Para ello se empleará la librería *Keras*, que automatiza la implementación y personalización de redes neuronales, y que además trabaja sobre la librería *tensorflow* [Abadi *et al.*, 2016], que organiza todo el flujo de procesamiento de datos para que esté disponible para las unidades de procesamiento. Toda la información de la librería *Keras* se encuentra disponible en su sitio web de documentación <https://keras.io>.

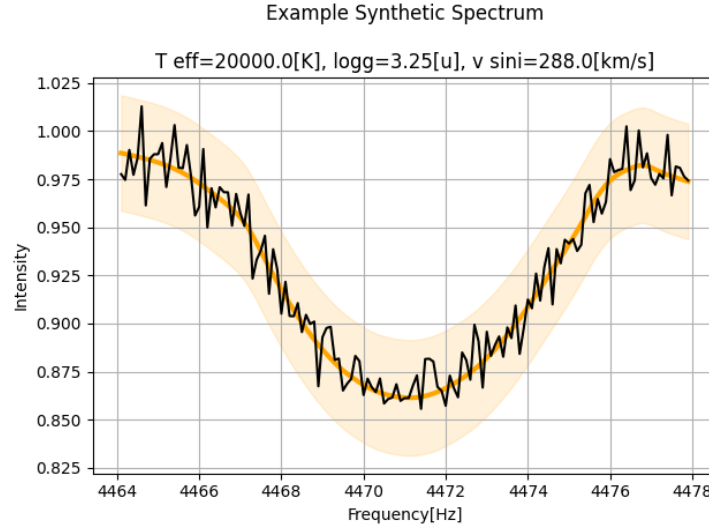


Figura 12: Ejemplo de espectro sintético con ruido añadido. La línea anaranjada corresponde a la expresión del espectro sintético teórico, sin alteraciones, mientras que la línea negra es el mismo espectro pero con ruido añadido. El área anaranjada corresponde al área donde se espera caigan el 99,7 % de los puntos con ruido, dada la distribución normal de este.

3.3. Red Neuronal Convolutional Canónica

La implementación de una CNN canónica tiene como finalidad, el poner un punto de partida para el análisis de espectros y predicción de parámetros estelares. Evaluar su eficacia y resultados, para determinar la factibilidad de esta metodología.

De acuerdo a lo anterior, resulta pertinente evaluar esta red en diferentes ámbitos de la experimentación para comparar resultados. Entre ellos, se evaluará la efectividad de esta red para datos sintéticos con y sin ruido, datos acotados y de rango extendido, y también para datos observados. Desde esta perspectiva, se propone el uso de la CNN de la Figura 13 para los propósitos mencionados.

A partir de los resultados obtenidos con esta red neuronal, se podrán determinar los lineamientos para experimentos siguientes y la configuración de datos a usar. En consecuencia de lo anterior, se podrá determinar bajo qué premisa se realizará la optimización de modelos con *Autokeras*.

3.4. Encontrar el modelo de predicción óptimo con Autokeras

Después de comprobar la efectividad de la metodología en la predicción de parámetros estelares, viene un proceso de optimización de la CNN. Esto con el objetivo de ampliar el espacio de búsqueda y encontrar la CNN con mejor desempeño sobre los datos.

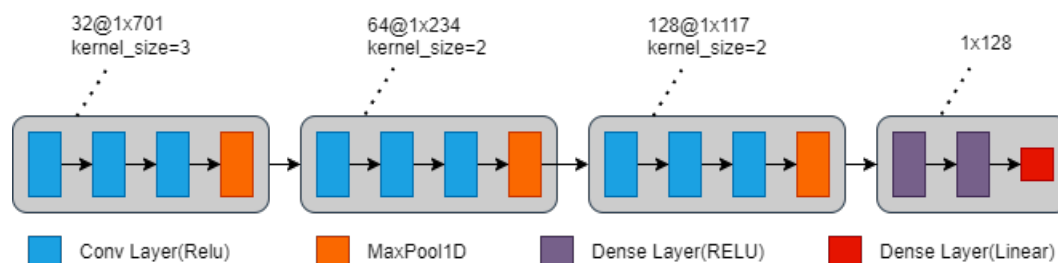


Figura 13: Arquitectura de CNN Canónica. La red se compone de tres bloques convolucionales con tres capas convolucionales antes de una capa *MaxPool* en cada uno de ellos. El primer bloque tiene tamaño de *kernel* igual 3, y los demás igual a 2. Cada bloque tiene 32, 64 y 128 filtros respectivamente. El último, es un bloque denso con dos capas densas de 128 filtros antes de la capa de regresión. Todas las capas convolucionales y densas tienen función de activación RELU a excepción de la última, que tiene función de activación lineal.

Para ello se utilizará la herramienta de aprendizaje automático llamada *Autokeras* [Jin *et al.*, 2019], que realizará una búsqueda sobre un espacio de hiperparámetros provistos, para encontrar aquellos que generen el mejor desempeño de la CNN. La metodología de búsqueda, a grandes rasgos se describe en dos partes:

1. **Bayesian:** La primera parte de la búsqueda, consiste en una metodología de optimización Bayesiana [Garnett, 2023] para encontrar el modelo de mejor desempeño dentro de un espacio de búsqueda “amplio”. Este proceso de optimización Bayesiana considera como parámetro de optimización el desempeño de la red en la métrica MSE para el conjunto de validación, y como variable independiente, a cualquiera de los parámetros provistos dentro del espacio de búsqueda. El algoritmo le otorga, en cada paso de ejecución, valores a las variables dependientes que podrían, potencialmente, arrojar el mejor desempeño. Para esta primera parte, dentro del espacio de búsqueda se consideraron un total de 500 modelos (intentos) de CNN diferentes. El espacio de búsqueda propuesto se encuentra detallado en el Anexo A.
2. **Greedy:** La segunda parte de la búsqueda, ya con información del desempeño de la red en un rango amplio, consiste en profundizar la búsqueda sobre una región de hiperparámetros focalizada en la región donde se encontró el mejor desempeño. Para ello, en esta parte se empleará un criterio avaro (*greedy*) sobre un espacio de búsqueda acotado y cercano a la región donde se encontró el mejor modelo de la parte anterior. Para esta segunda parte se consideraron un total de 300 modelos (intentos) de CNN diferentes. El espacio de búsqueda propuesto, para esta parte se encuentra detallado en el Anexo B.

3.5. Ensamble de modelos

Posterior a la selección de modelo, se realizará el entrenamiento de 100 modelos optimizados, con lo que será posible realizar el ensamble de modelos. Tomando en cuenta la información

entregada en la Sección 2.6 respecto a la metodología, y a la Sección 3.1.5 respecto de los conjuntos de datos disponibles, se describen los siguientes pasos para el ensamble:

1. Subdivisión del conjunto de datos original en conjuntos de entrenamiento, validación y pruebas.
2. Generación de 100 subconjuntos de entrenamiento a través de toma de muestras aleatorias con reemplazo (*bootstrapping*).
3. Asignación y entrenamiento de los modelos con subconjuntos de entrenamiento (*bagging*).
4. Cálculo de la métrica MAE de desempeño para cada modelo sobre el conjunto de validación.
5. Predicción de cada modelo sobre el conjunto de pruebas, o instancias observadas, dependiendo del caso.
6. Ensamble de modelos mediante promedio pesado de las predicciones (*weighted averaging*).
7. Cálculo del error asociado a cada predicción.

3.6. Predicción del ensamble sobre espectros observados

Para validar la efectividad del ensamble, se procederá a efectuar una serie de predicciones sobre espectros observados en los parámetros estelares mencionados anteriormente: T_{eff} , $v \sin i$ y $\log g$. En específico, las predicciones se efectuarán sobre los espectros de alta resolución publicados en el catálogo BeSOS, obtenidos a través del espectrógrafo PUCHE-ROS [Vanzi *et al.*, 2012] de estrellas tipo Be, cuyos parámetros fueron ya estudiados en [Arcos *et al.*, 2018] empleando metodologías tradicionales. El conjunto de espectros seleccionado corresponde a 68 espectros, y si bien el catálogo BeSOS contiene más espectros, sólo nos interesan aquellos para los cuales podamos comparar las predicciones asociadas con los valores determinados en [Arcos *et al.*, 2018].

Esta parte del trabajo nos dará una intuición de la efectividad de la metodología para la tarea de predicción de parámetros estelares, dado que las condiciones en las que se entrenarán los modelos corresponden a condiciones ideales, y no necesariamente coincidirán en cada caso con las mediciones efectuadas en la realidad. A partir de lo anterior se establecerán conclusiones acerca de qué se puede mejorar de la metodología en función de mejorar el desempeño del modelo y, finalmente, del ensamble de modelos.

3.7. Verificaciones sobre imprecisiones en el modelo y datos

Durante el proceso de experimentación, se verificó la efectividad de la metodología en cuestión, sin embargo, existieron factores relevantes que, se cree, pudieron afectar la precisión del modelo en algunos parámetros en particular.

Principalmente, el ruido adicionado, y el rango considerado para el espectro tuvieron influencia sobre las predicciones. Por tanto, se analizará en mayor profundidad aquellas configuraciones de datos que (en principio) no resultaban pertinentes al estudio, pero que resultaron importantes para validar la efectividad de las CNN en escenarios diferentes a los planteados inicialmente. Esto con el fin de entregar luces sobre aquellos factores que generan imprecisiones en la estimación. De acuerdo a esto, se mostrarán los resultados de exploración, realizada con posterioridad a los experimentos propuestos, con miras a dilucidar e interpretar errores en las predicciones del modelo de CNN optimizado.

En este sentido, resulta relevante definir una métrica frecuentemente usada en estudios astronómicos para describir la distancia entre un espectro observado y un espectro sintético que se desea ajustar. Definimos χ^2 como:

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - s_i)^2}{s_i} \quad (16)$$

Donde o_i corresponde a la intensidad registrada para un espectro observado en la frecuencia con índice i , análogamente para s_i , con n igual al número total de frecuencias arrojadas en la medición del espectro.

Finalmente, se comprobará la pertinencia de la metodología empleada comparando los espectros sintéticos con los espectros observados, esto contrastado con las predicciones del ensamble, para además evaluar el comportamiento y efectividad de la CNN optimizada en diferentes casos.

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

Para la siguiente sección se verificará la efectividad y se analizarán los resultados de *deep ensemble* en base al flujo de trabajo propuesto para la estimación de parámetros estelares.

En primera instancia se revisará una versión preliminar de entrenamiento y estimación para una CNN canónica con datos no acotados (ver Sección 3.1.5), tanto para los espectros sintéticos como observados y se evaluarán sus resultados. Después se comprobará el efecto que tiene entrenar una red canónica con datos acotados a la línea de emisión HeI.

Posteriormente, se estudiará la efectividad de una red neuronal optimizada usando *Autokeras* como herramienta de automatización de pruebas, con la idea de encontrar una CNN con un desempeño superior a las anteriores sobre los datos acotados.

Finalmente, con esta red neuronal optimizada, se discutirá la efectividad del ensamble de modelos respecto de los datos utilizados y se presentarán los resultados de las nuevas estimaciones para los espectros observados de estrellas tipo Be medidos con el espectrógrafo PUCHEROS con su respectiva incertidumbre.

Para facilitar el seguimiento de los experimentos efectuados, se dispone la Tabla 3 detallando la configuración de cada uno:

Sec.	Tipo CNN	Rango de Datos	Ruido Añadido
4.1	Canónica	No Acotados	No
4.2	Canónica	Acotados	No
4.3	Optimizada (Única)	Acotados	Sí
4.4	Optimizada (Ensamble)	Acotados	Sí
4.6.1	Optimizada	No Acotados	No
4.6.2	Optimizada	No Acotados	Sí

Tabla 3: Resumen de configuración de experimentos.

4.1. Red canónica sobre datos no acotados

La primera parte del trabajo a revisar, consiste en la exploración inicial sobre la efectividad de las CNN para la estimación de parámetros estelares. Para ello se llevó a cabo el entrenamiento de una red neuronal canónica con datos sintéticos sin acotar, es decir, con rango entre los 4440[Å] a 4510[Å], sin ruido añadido y con datos escalados entre el valor máximo y mínimo de cada frecuencia. La red canónica presenta las características indicadas en la Sección 3.3. Para esta arquitectura, se puede evidenciar la convergencia de la red en la Figura 14.

Los resultados de las métricas de desempeño de esta CNN se verifican en la Tabla 4. Para la función de pérdida (*loss*), se reporta un valor promedio de 0.0003, mientras que para MAE, el valor promedio es 0.0140. Estos valores (principalmente MAE) nos servirán de referencia respecto a optimizaciones evidenciadas en las secciones siguientes.

	T_{eff}^*	$\log g^*$	$v \sin i^*$	Overall
<i>loss</i>	0.0004	0.0003	0.0002	0.0003
MAE	0.0160	0.0136	0.0123	0.0140

Tabla 4: Métricas de desempeño, CNN canónica sobre datos sin ruido añadido y de rango extendido. Todas las métricas están calculadas sobre datos escalados.

Por otro lado, también resulta interesante revisar el desempeño de la CNN en cada uno de los parámetros estelares. Para la Tabla 4, se verifica que el parámetro de menor desempeño es el de T_{eff} , con MAE 0.0160, seguido de $\log g$, con MAE 0.0136, y por último $v \sin i$, con MAE 0.0123, que tuvo el mejor desempeño.

Los resultados de entrenamiento y validación se pueden visualizar en las Figuras 15 y 16, la primera figura corresponde la expresión la función de densidad de probabilidad (PDF) de los parámetros estelares de prueba, y de las predicciones realizadas por la red sobre los espectros correspondientes, diferenciadas color en rojo y verde respectivamente. La segunda figura corresponde a una visualización que permite comparar la distancia y el comportamiento entre las predicciones y los valores esperados.

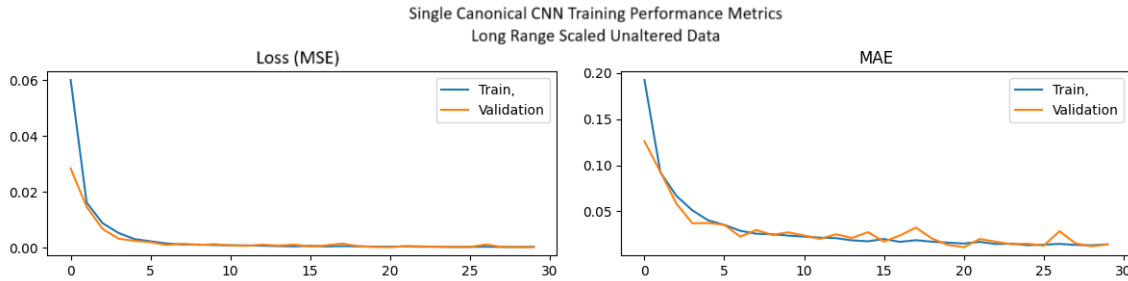


Figura 14: Métricas de Entrenamiento, CNN Canónica. En la figura se comprueba la disminución de los valores tanto de la función de pérdida como para MAE conforme se realizan más épocas de entrenamiento. Lo que corrobora la convergencia de la red convolucional.

En la Figura 15 se puede evidenciar que las PDF convergen en cada parámetro. La similitud de la curva para el parámetro $\log g$ es bastante favorable, sin embargo, todavía existe marco de mejora para la predicción de T_{eff} . Por otro lado, para el caso de $v \sin i$, si bien la gráfica no es favorable, se puede evidenciar del eje vertical que el rango de la densidad es bastante acotado, lo cual sigue un comportamiento esperado, dado que la distribución de probabilidad para dicho parámetro es uniforme y la variación corresponde al remuestreo de los datos para el entrenamiento y validación. En consecuencia, una pequeña variación en la muestra de los datos de predicción puede significar un gran cambio en la gráfica de la PDF, por lo que, el análisis del ajuste de la red en relación a este parámetro debe hacerse en conjunto con una

comparación de la distancia de cada predicción sobre su valor esperado, como en la Figura 16, para su correcta evaluación.

Con respecto de la Figura 16, en particular, para T_{eff} , existe una mínima tendencia a los valores por encima del valor esperado, y una ligera dispersión en valores esperados altos en comparación al resto. Para $\log g$ no existen mayores observaciones, lo que corrobora el buen ajuste evidenciado en la gráfica anterior. Por último, para $v \sin i$, si bien el ajuste fué desfavorable en la gráfica anterior, se puede evidenciar que los valores predichos son cercanos a los valores esperados, salvo una dispersión en los valores más altos y más bajos de este parámetro.

De esta manera se evidencia la factibilidad de estimación de los parámetros estelares al menos en condiciones ideales como las que se tiene trabajando con espectros sintéticos.

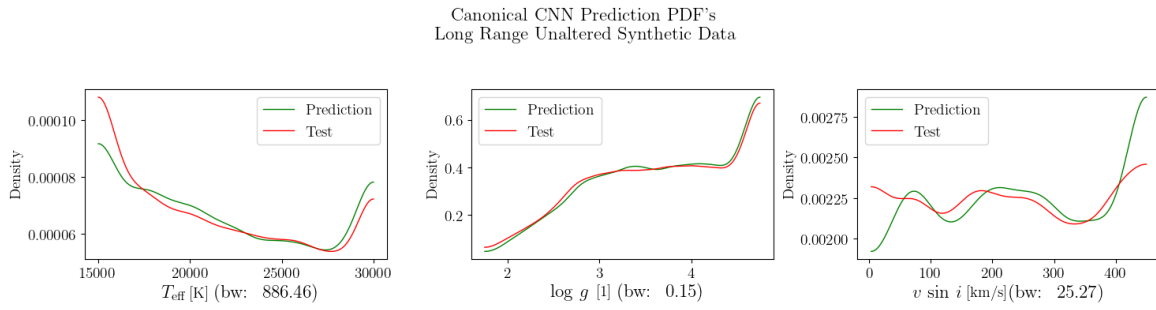


Figura 15: PDF's para predicción de CNN Canónica (en verde) y datos de prueba (en rojo): Datos sin ruido añadido y de rango extendido.

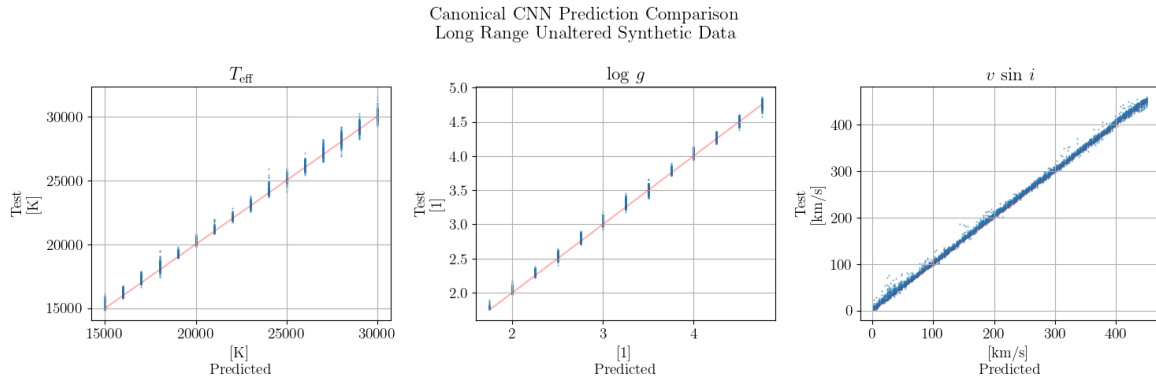


Figura 16: Comparación para predicción de CNN Canónica: Datos sin ruido añadido y de rango extendido.

Así, se corrobora una efectividad, en términos relativos dentro de los rangos previsto, aceptable para esta CNN. En condiciones ideales como en las que trabaja esta red, se aprecia un desempeño deseable de repetir en otras instancias.

4.2. Red canónica sobre datos acotados

A continuación se revisa la efectividad de la CNN canónica sobre datos acotados. En la Tabla 5 se tiene que el valor de la $loss$ es de 0.0011, y el de MAE, es de 0.0208, lo cual significa una disminución del desempeño respecto a los resultados del experimento anterior. Además, vemos que el parámetro que presenta un mayor decremento en su precisión, en ambas métricas, corresponde al de T_{eff} , seguido de $\log g$, mientras que $v \sin i$ no presenta cambios en el valor de alguna de las métricas.

En la Figura 17 se aprecia la PDF para las predicciones realizadas sobre datos acotados escalados entre máximo y mínimo para cada frecuencia. Visualmente se puede apreciar que para el parámetro $\log g$, el ajuste es similar a la CNN de la sección anterior. Sin embargo, existe un empeoramiento respecto al ajuste para la temperatura, dado que existe una oscilación entre los 18000[K] y los 26000[K]. Esta observación se confirma en la gráfica de la Figura 18, donde las predicciones en el rango de los 18000[K] y 22000[K] tienden ser mayores que el valor esperado, por tanto, y presumiblemente, esto aumentaría la densidad de probabilidad entre los 22000[K] y 26000[K] explicando de cierta manera los resultados de la gráfica anterior.

	T_{eff}^*	$\log g^*$	$v \sin i^*$	Overall
loss	0.0027	0.0005	0.0002	0.0011
MAE	0.0337	0.0171	0.0123	0.0208

Tabla 5: Métricas de desempeño, CNN canónica sobre datos sin ruido añadido y acotados. Todas las métricas están calculadas sobre datos escalados.

Con respecto a $v \sin i$, si bien la gráfica de la PDF en la Figura 17 no es favorable, se ratifica la aseveración del punto anterior en la Figura 18, donde se puede apreciar que, de todas maneras, y si bien existe un mayor grado de dispersión en los datos, el ajuste para la predicción de este parámetro sigue siendo aceptable.

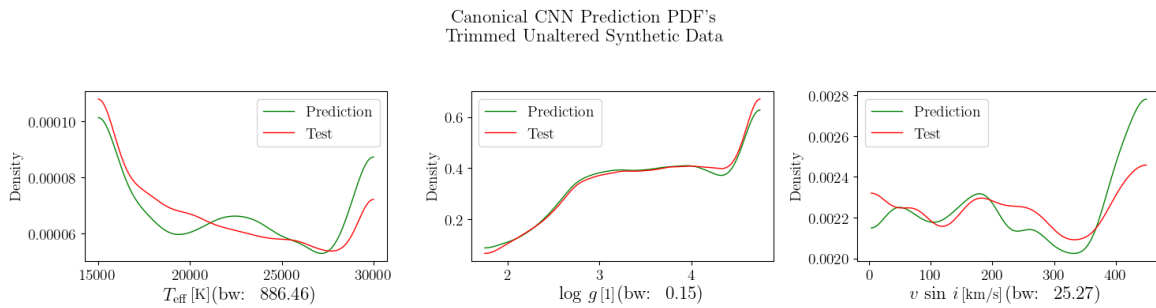


Figura 17: PDF's para predicción de CNN Canónica (en verde) y datos de prueba (en rojo): Datos escalados, sin ruido añadido y de rango acotado.

De esta manera se corrobora que la red tuvo menor desempeño con respecto a la configuración de datos anteriores, en los parámetros de T_{eff} y $\log g$. Sin embargo, no existe forma de encontrar y corroborar una red neuronal equivalente para la diferente cantidad de datos

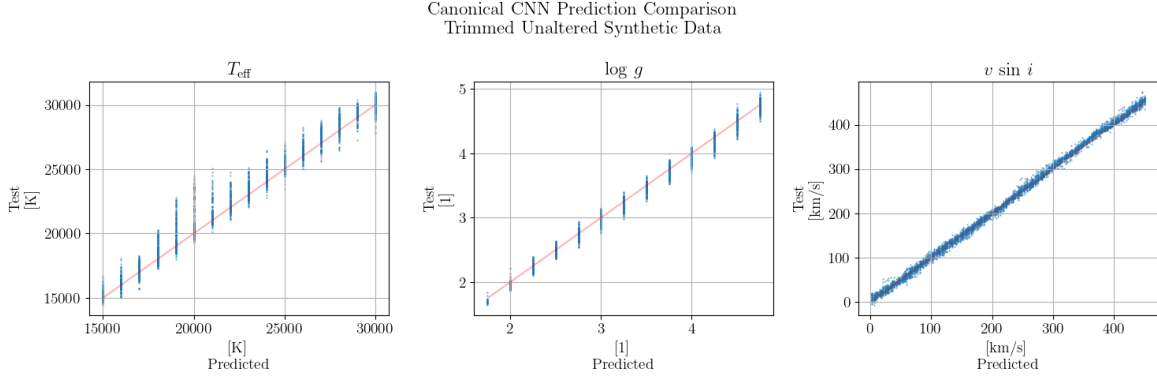


Figura 18: Comparación para predicción de CNN Canónica: Datos escalados, sin ruido añadido y de rango acotado.

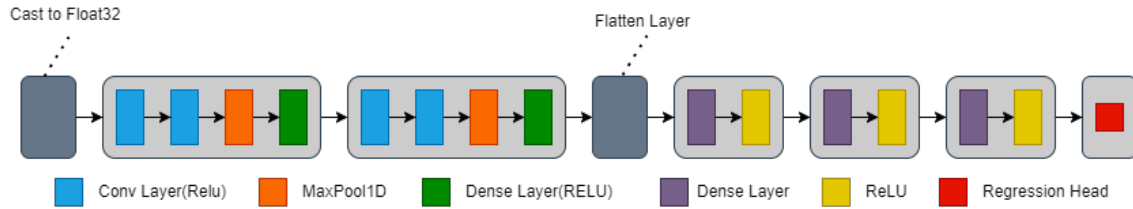


Figura 19: Representación de CNN Optimizada para datos acotados.

por espectro en ambos experimentos. Es por ello que resulta esencial revisar los resultados del proceso de optimización ya que, si bien no es posible homologar una red neuronal para ambas situaciones, esto permitirá dilucidar si es posible repetir o mejorar los resultados de la CNN canónica en el experimento anterior.

4.3. Optimización con Autokeras sobre datos acotados

En la segunda parte de este trabajo se revisarán los resultados de la CNN optimizada con la herramienta *Autokeras* realizando numerosas pruebas.

Los resultados sobre datos de prueba para la CNN optimizada están expresados en la Tabla 6. En esta ocasión, se tiene una disminución considerable del desempeño de la CNN para el parámetro de T_{eff} , con MAE 0.0656, así como de $\log g$, con MAE 0.0337. Por otra parte se aprecia mejora en el parámetro $v \sin i$ con MAE 0.0118.

	T_{eff}^*	$\log g^*$	$v \sin i^*$	Overall
loss	0.0099	0.0027	0.0003	0.0042
MAE	0.0656	0.0337	0.0118	0.0371

Tabla 6: Métricas de desempeño, CNN optimizada sobre datos sin ruido añadido y acotados. Todas las métricas están calculadas sobre datos escalados.

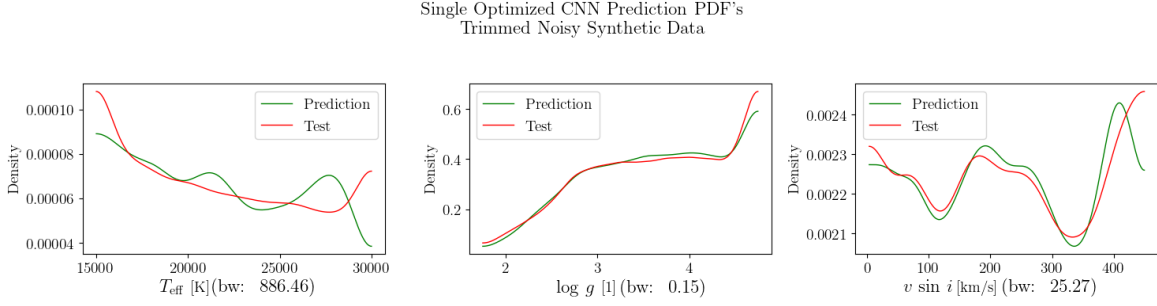


Figura 20: PDF's para predicción de CNN optimizada (en verde) y datos de prueba (en rojo): Datos ruidosos y de rango acotado.

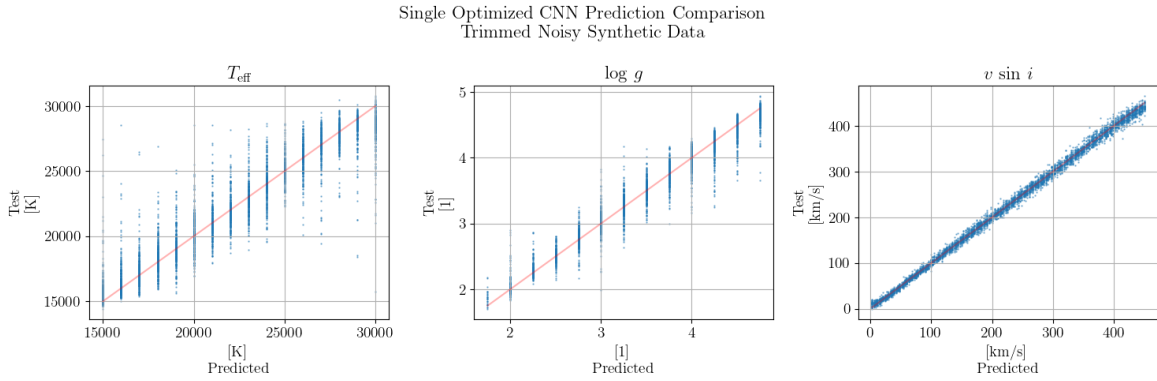


Figura 21: Comparación para predicción de CNN optimizada: Datos ruidosos y de rango acotado.

En la Figura 20 se pueden apreciar comportamientos particulares de la CNN en algunos aspectos. Para T_{eff} se aprecia una densidad de probabilidad en el vecindario de los valores 21000[K] y 28000[K] mayor a la esperada, y una densidad menor a la esperada en los valores más bajos, cercanos a 15000[k]. La gráfica de la Figura 21 en este parámetro permite corroborar una parte de estas observaciones, donde se observa una tendencia a valores mayores a los esperados en la parte baja de la gráfica, y una tendencia a valores menores a los esperados en la parte alta de la gráfica. Por lo demás, no es posible llegar a conclusiones respecto de la alta densidad en el vecindario de los 21000[K].

Con respecto a los parámetros $\log g$ y $v \sin i$, de la Figura 20 se puede evidenciar un ajuste deseable para la CNN optimizada única, sin embargo, en la gráfica de la Figura 20 se aprecia un aumento de la dispersión en relación a experimentos anteriores para todos los parámetros. Se hace necesario notar dificultades evidentes en la predicción de algunos casos de T_{eff} y $\log g$ con estimaciones alejadas del valor esperado en la recta. Para el $v \sin i$, en la misma figura, existe un muy buen ajuste para valores bajos de este parámetro, y existe una disminución de la precisión a medida que aumenta este valor.

4.4. Ensamble de modelos optimizados

Los últimos resultados a analizar, corresponden a los del ensamble de modelos optimizados. La arquitectura de la mejor CNN encontrada se encuentra en el Anexo D. En la Tabla 7 se muestran las métricas de desempeño para esta metodología. Comparablemente, existe una mejora en todas las métricas en relación al modelo optimizado único. Por otra parte, todavía existe un margen de mejora respecto a los datos de prueba.

	T_{eff}^*	$\log g^*$	$v \sin i^*$	Overall
loss	0.0094	0.0024	0.0003	0.0040
MAE	0.0620	0.0314	0.0100	0.0345

Tabla 7: Métricas de desempeño, ensamble de modelos optimizados sobre datos con ruido añadido y acotados. Todas las métricas están calculadas sobre datos escalados.

En la Figura 22 se pueden ver dos gráficas por parámetro a predecir. Las gráficas representan la PDF de 100 CNN de igual arquitectura que la red optimizada, entrenadas según el flujo descrito en la Sección 2.6. A partir de las gráficas en la parte superior de la figura se puede evidenciar que las diferentes CNN tienden a repetir el comportamiento de la red optimizada única, lo cual tiene un lado positivo, dado que las CNN convergen a un mismo patrón, y otro lado negativo, dado que repiten aquellas características desfavorables de la red anterior.

En la parte baja de la figura, se puede apreciar en el área sombreada de gráficas, el intervalo de confianza del 95 % para los tres parámetros. La mayor convergencia se aprecia para $\log g$, mientras que la mayor variabilidad se aprecia en el parámetro de $v \sin i$. Además de lo anterior, en particular para T_{eff} se logra apreciar que se replican los mismos desajustes en regiones evidenciadas en la CNN optimizada única.

En cuanto a la Figura 23, se puede apreciar un comportamiento similar al descrito para la CNN optimizada única. Se confirman las dificultades, en algunos casos, de la red neuronal para predecir valores de T_{eff} y $\log g$, con predicciones bastante alejadas del valor esperado. Por otra parte, se repite el aumento de imprecisión a medida que aumenta el valor del parámetro $v \sin i$.

Estos resultados, tanto para la CNN optimizada única como para el ensamble de modelos, son interpretables desde el punto de vista de los datos. Es esperable que el desempeño sea menor a los ejercicios anteriores, dado el ruido adicionado a los datos, lo que podría invisibilizar ciertas características que antes le permitían a la red neuronal realizar predicciones con mayor precisión.

Además, y con respecto al error percibido en las instancias de prueba, en la Figura 24 es posible apreciar el comportamiento de la cuantificación de error en la predicción. Para T_{eff} se puede apreciar que la distribución del error tiende a valores más altos conforme aumenta el valor de este parámetro. Caso opuesto se tiene para $\log g$, donde los errores se reducen cada vez más a medida que avanzamos en el eje del parámetro. Para $v \sin i$, se constata que la

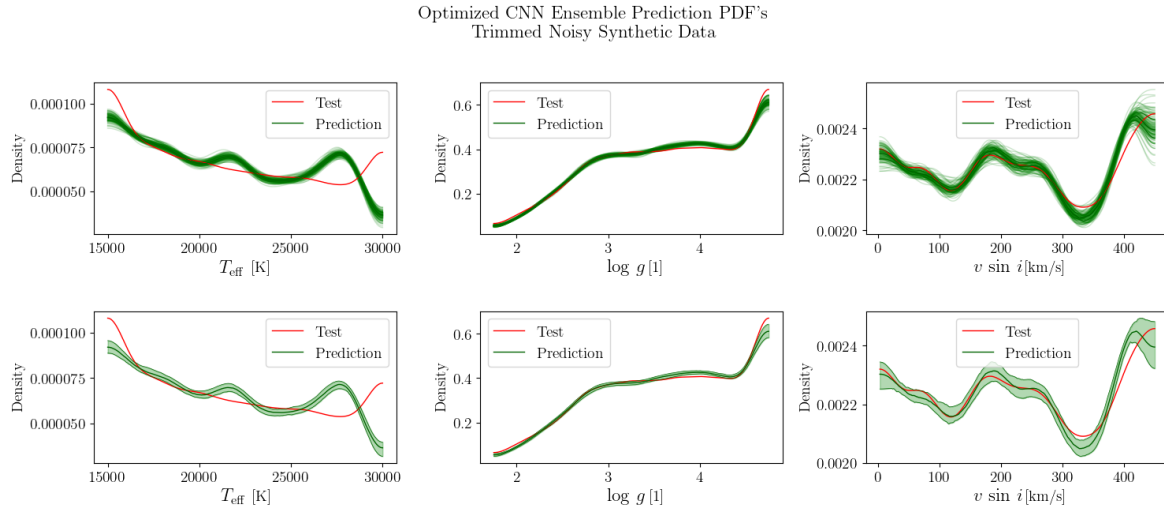


Figura 22: PDF's para predicción de ensemble de CNN optimizada: Datos ruidosos y de rango acotado.

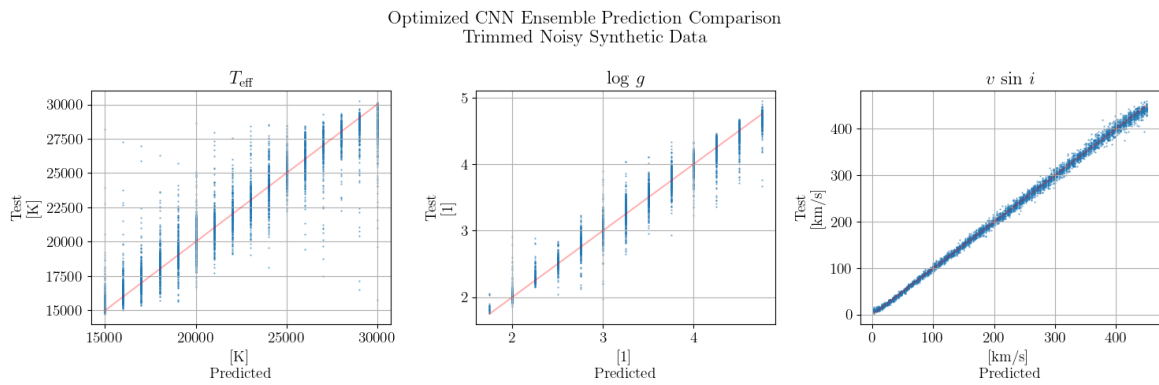


Figura 23: Comparación para predicción de ensemble de CNN optimizada: Datos ruidosos y de rango acotado.

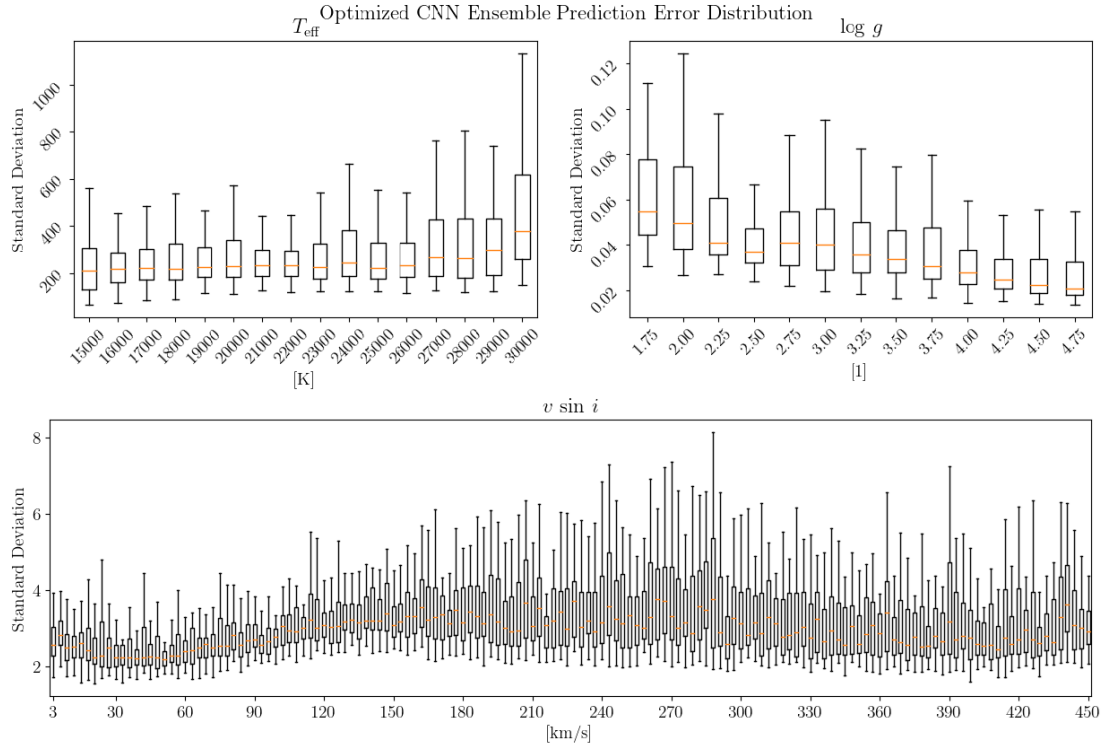


Figura 24: Tendencia de la desviación estándar en datos de prueba para el ensemble de modelos.

menor dispersión se encuentra en los valores más bajos, que es mayor en los rangos medios, y que vuelve a disminuir en rangos altos. A pesar de que, en general, el error es mayor en los rangos más bajos. Así, se pueden constatar que para dichas regiones del espacio de parámetros se tiene una mayor discrepancia entre los modelos y hay una mayor variabilidad en la predicción individual.

4.5. Predicción sobre espectros observados

Ahora, como se mencionó en la propuesta de trabajo, el objetivo final de este trabajo es efectuar predicciones sobre espectros observados, por lo que se presenta en el Anexo C, los resultados de la predicción de parámetros estelares sobre los espectros observados de estrellas con tipo espectral Be, del catálogo BeSOS. En la Figura 25, se puede apreciar las predicciones para dichos espectros, y en la Tabla 8 las métricas de desempeño respectivas. En general, respecto a la figura, es posible apreciar una dispersión significativa en la predicción con respecto al valor esperado en algunos parámetros. La mayoría de los resultados para T_{eff} se concentran en la parte baja de la gráfica, algunos bastante certeros en la predicción, mientras otros se encuentran bastante alejados de su valor esperado, tendencia que se repite a lo largo del rango de este parámetro. Para $\log g$ se puede apreciar una dispersión significativa de las predicciones con respecto a la recta, sin embargo, y a pesar de que la mayoría de los valores estimados están bajo la recta, se puede apreciar una tendencia de aumento del valor de la predicción cuando

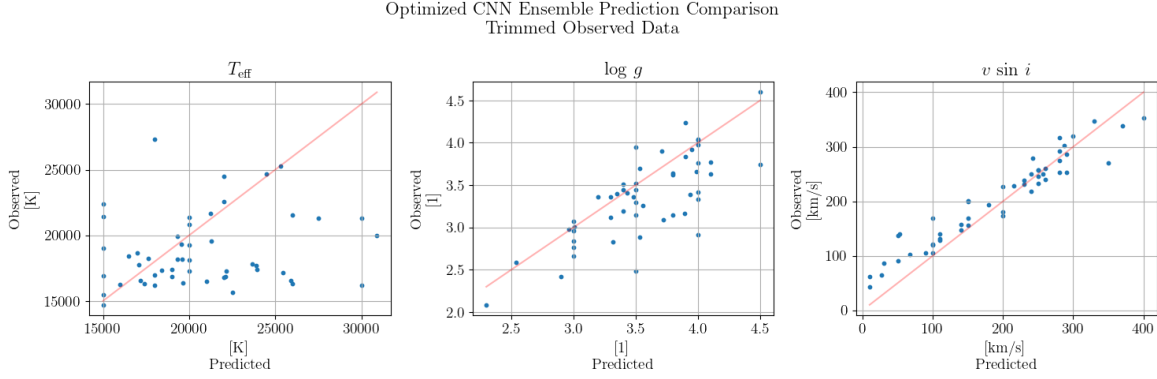


Figura 25: Comparación para predicción de ensemble de CNN optimizada: Datos observados de rango acotado.

	T_{eff}^*	$\log g^*$	$v \sin i^*$	Overall
loss	0.1063	0.0168	0.0058	0.0430
MAE	0.2361	0.0944	0.0593	0.1299

Tabla 8: Métricas de desempeño, ensemble de modelos optimizados sobre datos observados acotados. Todas las métricas están calculadas sobre datos escalados.

aumenta el valor esperado. En relación a $v \sin i$ se aprecia una gráfica favorable, donde los valores predichos se acercan bastante a la recta, si bien aún se aprecia un margen de mejora. Consecuentemente, lo expuesto se condice con lo con los resultados de la tabla de métricas, donde se observa un desempeño menor en relación a las instancias sintéticas para todos los parámetro.

4.6. Discusión sobre el tratamiento de los datos y la predicción

En base los resultados anteriores, resulta importante dilucidar las razones de las imprecisiones de ambos tipos de modelo para los datos tanto acotados, como de rango extendido. Esto permitirá conocer oportunidades de mejora para la metodología empleada e investigación, lo que puede ayudar a obtener mejores resultados a futuro para la tarea de predicción de parámetros estelares a partir de espectros observados.

En primera instancia, se tiene que la CNN canónica, tiene un margen de mejora considerable en cuanto a su ajuste, pero tiene buen desempeño en todos los parámetros, por lo que sería interesante saber, incluso si es en condiciones ideales, si es posible encontrar un modelo que ajuste perfectamente todos los parámetros.

Por otra parte, se aprecia en gran medida, dificultades de las CNN (incluso para la CNN optimizada), para estimar el valor esperado del parámetro T_{eff} al trabajar con datos acotados. Por lo que cabe preguntarse si los datos con rango extendido están mejor condicionados para la estimación de este parámetro.

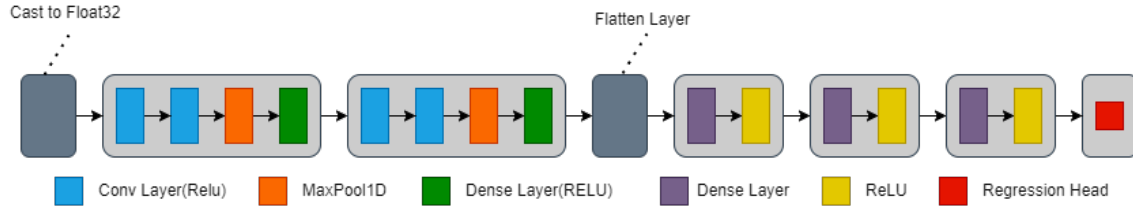


Figura 26: Representación de CNN Optimizada para datos de rango extendido.

Adicionalmente, si se tuviese una configuración de datos y CNN que resolviese las interrogantes anteriores, sería relevante estudiar qué sucede al entrenar y realizar predicciones con esta CNN con datos que tengan ruido agregado.

En base a estas ideas, se presentan algunos estudios que intentan arrojar luz a estas interrogantes. Primero, se presenta un estudio de CNN optimizada con *Autokeras* sobre una situación ideal, vale decir, datos de rango extendido y sin ruido, para resolver el problema del ajuste “perfecto”. Se demostrará que, al mismo tiempo, esta nueva CNN optimizada resuelve el problema de las imprecisiones en T_{eff} . Finalmente, se revisará qué sucede al entrenar esta red neuronal con datos que tienen ruido adicionado.

4.6.1. CNN óptima en condiciones ideales

Para esta parte, se llevó a cabo nuevamente una optimización con *Autokeras*, en un total de 50 intentos (la configuración para el entrenamiento y la mejor arquitectura de CNN hallada se encuentran en los Anexos E y F respectivamente). Se emplearon, también los datos de rango extendido escalados y sin ruido adicional, lo que constituye condiciones ideales para la ejecución de la CNN. En base a esto, se obtuvieron los resultados de la Tabla 9.

	T_{eff}^*	$\log g^*$	$v \sin i^*$	Overall
loss	4e-05	4e-05	2e-05	3e-05
MAE	0.0045	0.0046	0.0033	0.0041

Tabla 9: Métricas de desempeño, CNN optimizada sobre datos sin ruido añadido y de rango extendido.

En la Tabla 9, se encuentra el mejor resultado de este estudio con respecto a la estimación de parámetros, al menos desde la perspectiva de los datos de prueba, donde la *loss* llega a un valor de lo 3e-05, y MAE alcanza un valor de 0.0041, lo que se puede traducir en un error promedio de menos del 0.5 %.

En las Figuras 27 y 28 se puede comprobar visualmente el ajuste de esta CNN optimizada, donde podemos ver, en la primera gráfica, que la PDF de los datos de predicción se ajusta casi perfectamente en todos los parámetros. También en la segunda gráfica mencionada, se puede ver una dispersión mínima de las predicciones respecto del valor esperado.

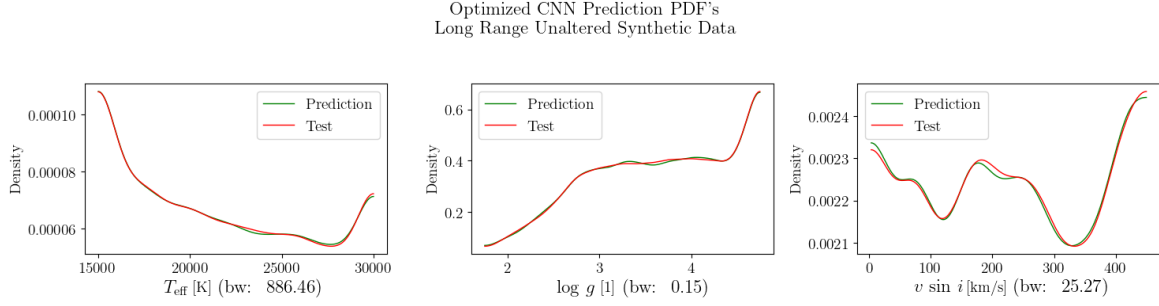


Figura 27: PDF's para predicción de CNN optimizada (en verde) y datos de prueba (en rojo): Datos sin ruido añadido y de rango extendido.

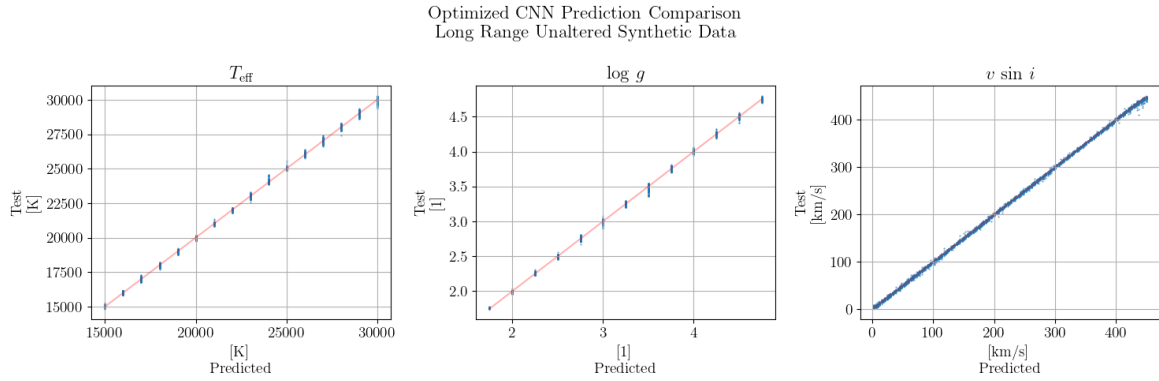


Figura 28: Comparación para predicción de CNN optimizada: Datos sin ruido añadido y de rango extendido.

Ante la evidencia anterior, se concluye sí es posible obtener una CNN con alto desempeño en condiciones ideales. Además de lo anterior, se puede interpretar que el rango del espectro que la CNN es capaz de percibir, tiene una influencia positiva en cuanto a la precisión de la misma, esto para todos los parámetros, pero sobre todo se aprecia una mejora considerable para T_{eff} , por lo que también se cierra la interrogante acerca de si es posible mejorar la predicción de este parámetro al incluir más datos.

Por otra parte, todavía queda una interrogante relevante, ya que si bien los resultados de esta CNN son prometedores, es necesario recordar que ésta fue entrenada bajo condiciones ideales que no necesariamente reflejan la realidad de las mediciones tomadas mediante un espectrógrafo, que pueden tener ruido no despreciable. Ante esta situación expuesto, es necesario validar la calidad de la CNN anterior, esta vez sobre datos con ruido adicionado.

4.6.2. Ruido en en los datos

Para el entrenamiento de la misma CNN del punto anterior, esta vez, con datos con ruido adicionado descrito en la Sección 3.1.5, se reportan los resultados en la Tabla 10:

En comparación al experimento anterior, se tiene una disminución en cuanto a la precisión

	T_{eff}^*	$\log g^*$	$v \sin i^*$	Overall
loss	0.0014	0.0006	0.0002	0.0007
MAE	0.0260	0.0181	0.0116	0.0186

Tabla 10: Métricas de desempeño, CNN optimizada sobre datos con ruido adicionado y de rango extendido. Todas las métricas están calculadas sobre datos escalados.

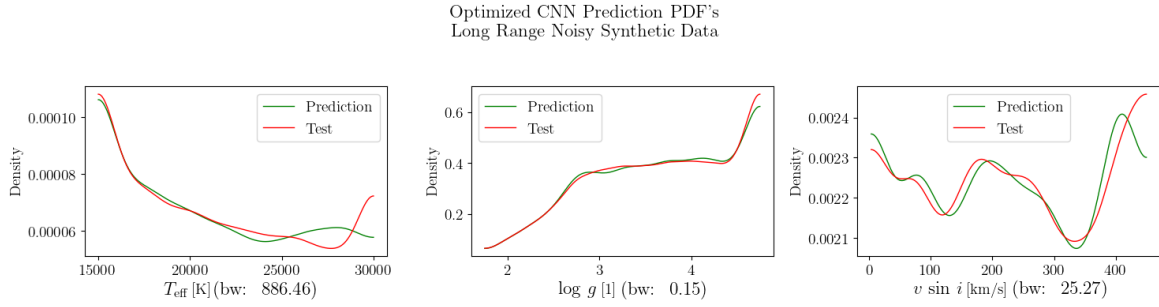


Figura 29: PDF's para predicción de CNN optimizada (en verde) y datos de prueba (en rojo): Datos con ruido y de rango extendido.

de la CNN. Este comportamiento es esperable a la adición de ruido. Sin embargo, esta CNN todavía sostiene un mejor desempeño que la red optimizada para datos acotados.

En la Figura 29 se puede ver que el ajuste para las PDF también disminuyó. Para el parámetro T_{eff} se observa un desajuste a partir de los 22000[K], sin embargo en la gráfica comparativa de la Figura 30, no se logra apreciar con claridad la razón de este comportamiento. En $v \sin i$ se tiene que existen varias regiones donde hay desajustes al mismo tiempo. También, una situación interesante se puede comprobar en este caso para $\log g$, donde en la Figura 29 se puede ver que la PDF no revela mayor cambio en el ajuste, sin embargo, al mirar la gráfica del mismo parámetro en la Figura 30, se puede apreciar una mayor dispersión con respecto al experimento anterior, lo que indica que a pesar del aparente buen ajuste de los datos en este parámetro en la primera figura, esta no necesariamente representa la realidad de una buena predicción.

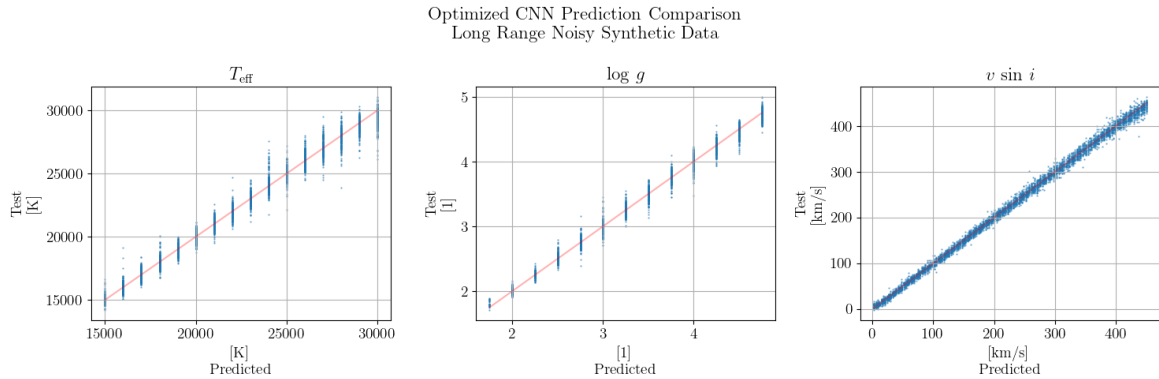


Figura 30: Comparación para predicción de CNN optimizada: Datos con ruido y de rango extendido.

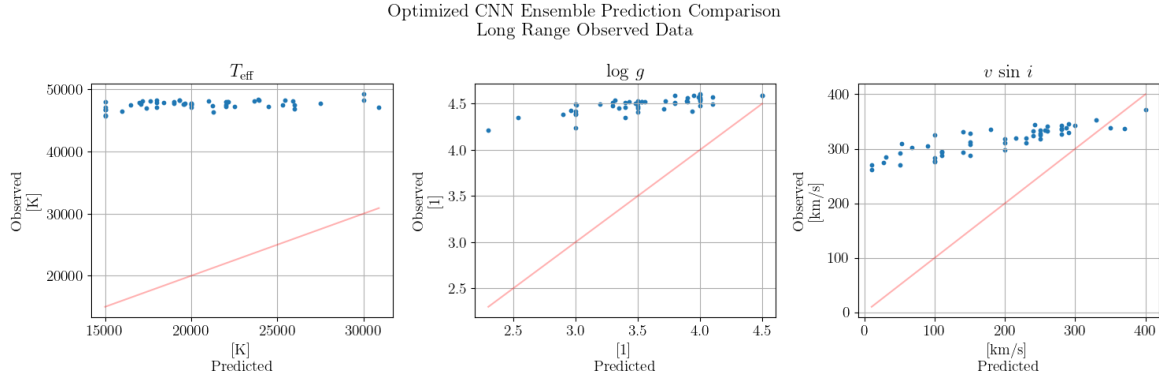


Figura 31: Comparación para predicción de CNN optimizada: Datos observados de rango extendido.

Respecto a este experimento de validación, se puede concluir que, como es esperable, el ruido tiene un efecto significativo en el ajuste de la CNN, pero sobre todo tiene un efecto considerable en cuanto a la predicción del parámetro T_{eff} , lo que podrá dar pistas de los pasos a seguir para mejorar este aspecto en la predicción de parámetros estelares para espectros observados, que conllevan ruido inherente al proceso de medición.

Por otra parte, si bien los resultados para esta disposición de datos y entrenamiento son alentadores, en la Figura 31 se puede observar el sobre-ajuste de la CNN con respecto a los datos observados, donde se aprecia que las predicciones se encuentran, para cada parámetro, muy por encima del valor esperado, y en un rango acotado (considerando la gama de valores posibles de cada parámetro), hecho que es indicador de una mala generalización de la red.

4.6.3. Discusión sobre la predicción en espectros observados de rango extendido.

En el Anexo C es posible encontrar los resultados de las predicciones efectuadas sobre espectros observados. Como se puede comprobar, si bien existen predicciones bastante certeras en algunos espectros, con errores relativos menores o cercanos al 5 %, también existen predicciones que escapan bastante del valor real, con estimaciones que superan el 40 % de error relativo.

En base a esta observación, resulta importante dilucidar el motivo de estos resultados. Como se ha visto con anterioridad, la CNN es sensible a la calidad de los datos, por lo que una comparación entre los espectros sintéticos (espectros con los que fué entrenada la CNN) y los espectros observados podría ser provechosa para conocer el comportamiento de la CNN. Para tener una idea, en la Figura 32 se muestra un caso ideal para la predicción para la CNN, y en la Tabla 11 se encuentran los respectivos parámetros de cada espectro.

En la gráfica, se pueden visualizar tres espectros diferentes: el espectro observado de una

¹Para esta estrella, el error asociado a la predicción de $\log g$ es menor a la mínima unidad de error posible, por lo que no se reporta.

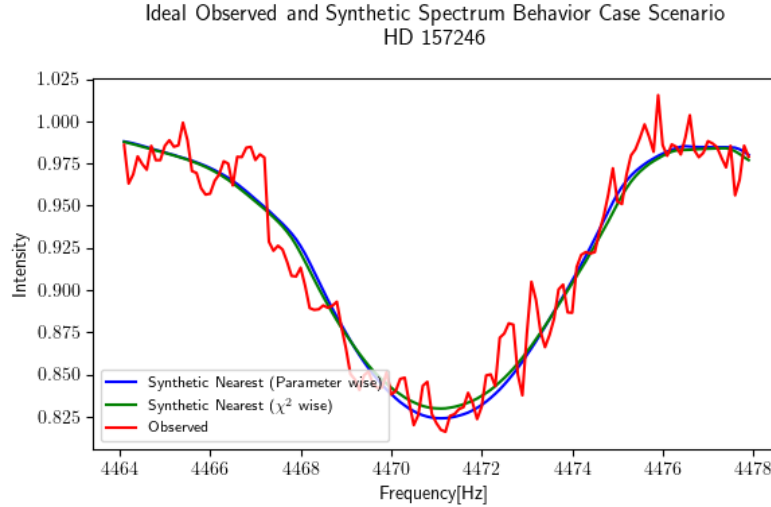


Figura 32: Caso ideal de comparación entre espectro observado y sintéticos: HD157246.

Star	T_{eff} [K]	$\log g$ [1]	$v \sin i$ [km/s]
HD157246			
Literature [Arcos <i>et al.</i> , 2018]	21240	3.5	230
Synthetic Nearest (Parameter wise)	21000	3.5	231
Synthetic Nearest (χ^2 wise)	22000	3.5	243
Prediction	21646 ± 304	3.4^1	239 ± 6

Tabla 11: Tabla comparativa espectro observado vs sintético: HD157246.

estrella (en rojo): el espectro sintético (extraído del conjunto de espectros generados) con los parámetros más cercanos al espectro observado (en azul); y el espectro sintético más cercano al espectro observado en términos de χ^2 (en verde). Cabe mencionar que los dos últimos espectros fueron representados sin ruido añadido para una fácil interpretación. Los parámetros correspondientes a cada espectro se describen en la Tabla 11 además de la predicción arrojada por el ensamble de CNN.

La Figura 32 representa lo que sería deseable encontrar respecto de un espectro observado de una estrella. El espectro sintético más cercano en cuanto a χ^2 describe fielmente la tendencia de la forma del espectro observado. En consecuencia, el caso no es extraño para la red neuronal y la predicción resulta certera, como lo indica la Tabla 11, donde el error promedio para los tres parámetros es de un 2.2 %.

Ahora bien esto no explica las situaciones donde sí se producen errores. Para ello hace falta adentrarse en estas situaciones en específico. A continuación se presenta la Figura 33 con tres gráficas comparativas de espectros observados y sintéticos. Cada gráfica se corresponde con una entrada de la Tabla 12 que describe los parámetros de cada espectro para cada estrella de la figura.

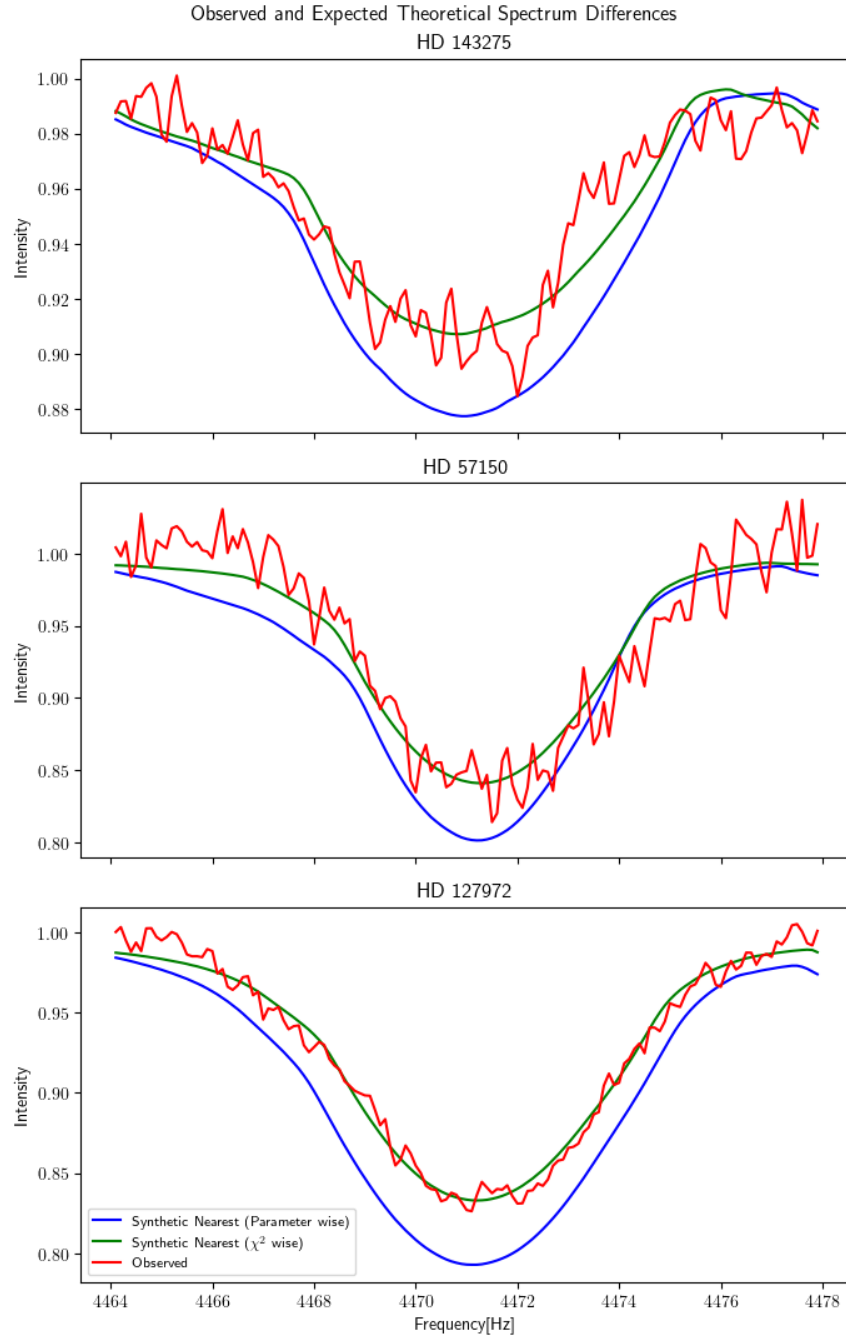


Figura 33: Diferencias entre espectros observados y teóricos. HD75311, HD57150 y HD1279972.

De las tres gráficas de la 33, se observa que el espectro más cercano en términos de χ^2 , no necesariamente coincide con aquel que es cercano en términos paramétricos, lo que da cuenta de diferencias significativas entre los espectros con los que se entrenó la CNN y aquellos espectros que se espera sea capaz de predecir. A demás de lo anterior, se podría considerar que el espectro más cercano en términos de χ^2 representa una curva suavizada del espectro observado, y que este último engloba al primero, pero incluso con esta idea en mente, existen

Star	T_{eff} [K]	$\log g$ [1]	$v \sin i$ [km/s]
HD143275			
Literature [Arcos <i>et al.</i> , 2018]	30900	3.5	257
Synthetic Nearest (Parameter wise)	30000	3.5	258
Synthetic Nearest (χ^2 wise)	20000	2.5	249
Prediction	19959 ± 522	2.5 ± 0.1	250 ± 13
HD57150			
Literature [Arcos <i>et al.</i> , 2018]	22000	3.5	180
Synthetic Nearest (Parameter wise)	22000	3.5	180
Synthetic Nearest (χ^2 wise)	16000	3.0	201
Prediction	16783 ± 1362	3.1 ± 0.2	194 ± 16
HD127972			
Literature [Arcos <i>et al.</i> , 2018]	21000	3.95	240
Synthetic Nearest (Parameter wise)	21000	4.0	240
Synthetic Nearest (χ^2 wise)	16000	4.0	219
Prediction	16471 ± 289	3.9 ± 0.2	219 ± 8

Tabla 12: Diferencias entre espectros observados y teóricos. Tabla de parámetros. HD143275, HD57150 y HD127972.

regiones donde el espectro observado no coincide con esta curva suavizada, por lo que es esperable que las predicciones sobre los espectros observados no sean certeras.

Adicionalmente, observando las entradas correspondientes a cada gráfica en la Tabla 12, se puede observar que en estos casos, donde existe una diferencia apreciable entre el espectro observado y el sintético más cercano en términos de parámetros, la predicción arrojada es mucho más cercana a los parámetros del espectro cercano en χ^2 que al valor del espectro observado, observación de la cual se pueden obtener dos conclusiones:

- El funcionamiento de la CNN, en general, trabaja de manera similar a las metodologías que ya se usaban en este campo. Es decir, los valores arrojados por el ensamble se corresponden con los parámetros de la curva que mejor se ajusta, o que más se parece, al espectro observado.
- No es razonable, dadas las diferencias entre ambos tipos de datos, esperar que la CNN realice una predicción robusta en casos como estos. La distribución de los datos de entrenamiento es diferente a la de los datos que se desea estimar, por tanto, es esperable que la predicción no sea certera cuando el espectro observado no se parece en forma a su símil teórico.

Estas conclusiones ayudan a clarificar una parte de los errores en las predicciones. Sin embargo existen otros casos, que se revisarán en los puntos siguientes, que se pueden evidenciar

a manera de complementar lo que sabemos de los errores en la predicción.

En la Figura 34 se presentan dos gráficas que representan una situación mucho menos frecuente, pero que también puede producir error en la predicción. Lo que se aprecia en ambas figuras puede describirse como un desplazamiento o sesgo del valle que representa línea de absorción presente. En síntesis, lo que quiere decir es que el valor más bajo en cuanto a intensidad del espectro observado, no se encuentra en la misma frecuencia que el valor más bajo del espectro teórico. Por ende, la información que se le entrega a la CNN puede resultar imprecisa, y resultar en errores de predicción. Para este caso, los parámetros correspondientes se encuentran en la Tabla 34, y los errores de predicción en cuanto al parámetro $\log g$ para las estrellas HD56139 y HD167128 resultan en un 18.6 % y un 17.2 % respectivamente.

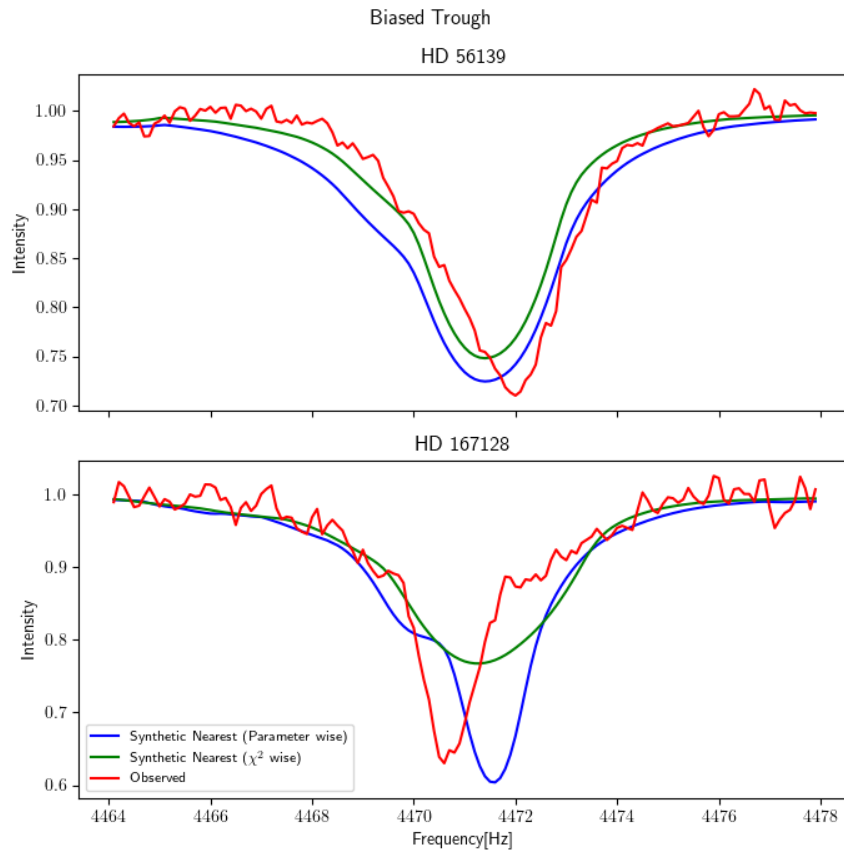


Figura 34: Desplazamiento de la línea de absorción. HD56139 y HD167128.

Por último otro tipo de análisis que se puede realizar, corresponde a comprobar la cantidad de ruido, vale decir dispersión asociada, que tienen las mediciones en particular. En la Figura 35 se muestra un caso con considerable dispersión de intensidad en el espectro observado de la estrella HD 209409. Para este espectro, no es relevante evaluar analíticamente la forma de alguna de las curvas representativas, sino, y dado el conocimiento previo respecto a escenarios favorables para el desempeño del ensamble de modelos, evaluar si es razonable esperar resultados positivos en casos como este. Los parámetros estimados para este espectro, descritos

Star	T_{eff} [K]	$\log g$ [1]	$v \sin i$ [km/s]
HD56139			
Literature [Arcos <i>et al.</i> , 2018]	17170	3.89	100
Synthetic Nearest (Parameter wise)	17000	4.0	99
Synthetic Nearest (χ^2 wise)	16000	3.25	96
Prediction	$16567 \pm 1305 \pm 522$	3.2 ± 0.1	106 ± 12
HD167128			
Literature [Arcos <i>et al.</i> , 2018]	20000	3.8	50
Synthetic Nearest (Parameter wise)	20000	3.75	51
Synthetic Nearest (χ^2 wise)	20000	3.25	132
Prediction	19265 ± 6602	3.1 ± 0.1	138 ± 10

Tabla 13: Desplazamiento de la línea de absorción/ Tabla de parámetros. HD56139 y HD 167128.

en la Tabla 14, junto con los parámetros de los espectros sintéticos, se encuentran en la Tabla 14, e indican un error para T_{eff} del 49.4 %, para $\log g$ del 5.4 %, y para $v \sin i$ del 22.7 %.

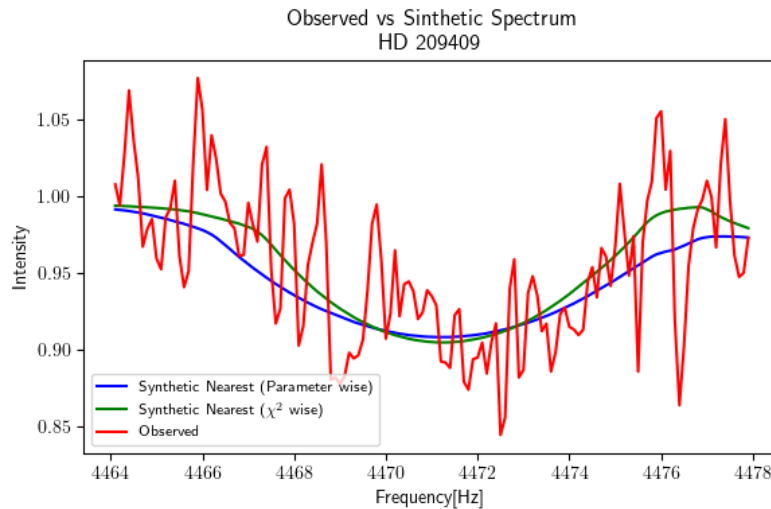


Figura 35: Ruido en la medición del espectro. HD209409.

Resumiendo los experimentos realizados, se adjunta, en la Tabla 15 un resumen de los resultados en la métrica MAE.

Star	T_{eff} [K]	$\log g$ [1]	$v \sin i$ [km/s]
HD209409			
Literature [Arcos <i>et al.</i> , 2018]	15000	3.0	350
Synthetic Nearest (Parameter wise)	15000	3.0	351
Synthetic Nearest (χ^2 wise)	15000	2.5	282
Prediction	22411 ± 4760	2.8 ± 0.3	271 ± 17

Tabla 14: Ruido en la medición del espectro. Tabla de parámetros. HD209409.

Sec.	Tipo CNN	Rango de Datos	Ruido Añadido	T_{eff}	$\log g$	$v \sin i$	General
4.1	Canónica	No Acotados	No	0.0160	0.0136	0.0123	0.0140
4.2	Canónica	Acotados	No	0.0337	0.0171	0.0123	0.0280
4.3	Optimizada (Single)	Acotados	Sí	0.0656	0.0337	0.0118	0.0371
4.4	Optimizada (Ensemble)	Acotados	Sí	0.0620	0.0134	0.0100	0.0345
4.4	Optimizada (Ensemble)	Acotados(Observados)	-	0.2361	0.0944	0.0593	0.1299
4.6.1	Optimizada	No Acotados	No	0.0045	0.0046	0.0033	0.0041
4.6.2	Optimizada	No Acotados	Sí	0.0260	0.0181	0.0116	0.0186

Tabla 15: Tabla resumen de resultados MAE. La columna “General” se constituye promediando los valores MAE de todos los parámetros.

CAPÍTULO 5

CONCLUSIONES

A partir de la experimentación realizada, se puede percibir un desempeño aceptable de la metodología de las CNN en general respecto de la tarea de predicción de parámetros estelares, llegando a un error menor al 1 % en el caso ideal, y del ensamble de modelos, llegando a un error menor al 5 % en los datos de prueba de espectros sintéticos. Ahora bien, como se evidenció en varias fases de los experimentos, el parámetro más regular en la predicción es el de $v \sin i$, sin embargo, todavía existe un marco de mejora considerable para los parámetros de T_{eff} y $\log g$. Por otra parte, el tratamiento de los datos resulta de gran relevancia para lograr predicciones más certeras.

El rendimiento más bajo en términos generales se dió en el parámetro T_{eff} para la mayoría de instancias, donde se aprecia la mayor dispersión y los mayores valores de las métricas desempeño. Por otra parte, se aprecia una mayor dificultad de los predictores, en este parámetro, cuando la cantidad de datos es está acotada de manera exclusiva a la línea de absorción HeI, lo que entrega indicios de posible mejores resultados, al tomar conscientemente un rango de espectro mayor para alimentar a la CNN. Trabajo futuro puede ser tomar más de una línea de emisión relevante en estrellas de tipo B al mismo tiempo, y entrenar la red neuronal con ellas. El parámetro $\log g$, tuvo un desempeño medio en relación a los demás ya que tuvo resultados prometedores en las instancias con datos sintéticos, si bien se evidencia que aún hay trabajo por hacer. Por otro lado, el parámetro $v \sin i$ fué el que tuvo mejor rendimiento en todas las pruebas, dado que se observa el menor error a lo largo de los experimentos.

Respecto del ensamble de modelos, se concluye que fue efectivo en mejorar la efectividad de las predicciones, donde se aprecia una mejora del 7 % respecto del modelo optimizado único en la métrica MAE para los espectros sintéticos. También, esta metodología entregó información relevante respecto a la dispersión en la predicción considerando los resultados de prueba con datos sintéticos, donde se puede identificar que los resultados con mayor desviación estándar se dieron para las instancias en las que se tiene menos datos disponibles, como valores bajos tanto para T_{eff} como $\log g$.

En relación la predicción de espectros observados, el error general llega al 13 % para la predicción con ensamble. Sobre este ítem, se corrobora el comportamiento ya mencionado anteriormente respecto al error en las predicciones de los parámetros. El parámetro con desempeño más bajo es T_{eff} , con un 24 % de error en la métrica MAE, seguido de un 9 % en $\log g$ y un 6 % en $v \sin i$ en la misma métrica. Estos valores sugieren un margen de mejora respecto del ajuste de la red neuronal, junto con aristas a abordar para mejorar la predicción de cada parámetro. En concordancia con lo anterior, el análisis de los factores de error en la predicción resulta pertinente para dilucidar metodologías para trabajar estas aristas.

Frente a los errores en la predicción de parámetros observados se lograron importantes descubrimientos. El comportamiento de la predicción de la CNN sigue de cerca la forma y ruido

de los espectros, y es sensible a los cambios que hay entre un espectro y otro. Por lo tanto, en varios casos, la diferencia entre la predicción y el valor esperado es atribuible a la medición del espectro y no al desempeño del predictor. Por otra parte, también se encontraron formas de ruido que no se habían considerado inicialmente. Entre ellas el desplazamiento respecto a la frecuencia esperada de la línea de absorción HeI. En consecuencia, podría llegar a ser interesante considerar otra forma de ruido adicional a los datos, que aborde este tipo de situación. También podría considerarse que el desplazamiento en cuanto a intensidad registrada en la línea de absorción puede ser una forma de ruido. Sin embargo, como la CNN es sensible a esta medida, el concepto tendría que abordarse con extremo cuidado antes de ser tratado.

Otro factor que incide en el error de predicción es el ruido inherente a la medición en cada punto, ya que el ruido presente en algunos espectros resulta mucho mayor que el esperado durante la fase de entrenamiento. Por lo tanto, una solución podría ser modelar adecuadamente el ruido presente en esta fase, de modo que los ejemplos vistos por la CNN se asemejen más a los que se encuentran en la realidad. Además, otra aproximación podría ser considerar diferentes niveles de ruido al construir el conjunto de entrenamiento, lo que permitiría abordar este aspecto del problema y, al mismo tiempo, generar más datos de entrenamiento para la CNN.

Otra temática con respecto al error en la predicción puede ser la cantidad de ejemplos disponibles, ya que el paso (*step*) de los parámetros en los modelos atmosféricos no permite generar una mayor cantidad de ejemplos. Si bien v sin i es un parámetro ajustable para la generación de espectros, el rango de los demás parámetros limita considerablemente la cantidad de espectros disponibles. Esto puede llevar a que la CNN entrenada no cuente con suficiente “experiencia” para ajustar correctamente todas las instancias.

De acuerdo a lo expresado, se puede decir que la metodología de ensamble es funcional, pero, de momento, todavía existen aspectos a mejorar en métodos que permitan abordar la predicción de los parámetros T_{eff} y $\log g$ con mayor certeza. Se mencionan específicamente estos parámetros, ya que la escala del error, en ciertas instancias, es considerable en términos astronómicos, donde una diferencia de 2000[K] en T_{eff} , como la desviación estándar correspondiente a algunas predicciones, puede llevar a la clasificación errónea de una estrella.

Además, en relación a los tipos de error descritos, y con respecto a la predicción del ensamble sobre espectros observados, cabe mencionar que el análisis sobre la información que entrega la desviación estándar pierde validez en comparación a la instancia con espectros sintéticos. Principalmente, por que se pueden observar predicciones con error mayor al 40 %, pero con desviación estándar menor al 5 %. Esto, presumiblemente, puede deberse a las fuentes de error identificadas anteriormente. Una explicación posible puede ser, por ejemplo, dado un caso de desplazamiento de la línea espectral observada respecto de la teórica, varios modelos realizan estimaciones similares, que estarán erradas, pero que tendrán baja desviación estándar debido a la similitud en las predicciones. En este sentido, y habiendo resuelto algunos problemas relacionados al desplazamiento de la expresión de la línea espectral, modelar el ruido en cuanto a intensidad percibida en la medición podría resultar más interesante.

En resumen, la metodología empleada en este trabajo arrojó resultados favorables desde el punto de vista del uso de los espectros sintéticos para la predicción de parámetros estelares. A su vez, se detectaron deficiencias en cuanto a la predicción de parámetros de espectros observados. De acuerdo a estos resultados, trabajo futuro con metodologías que aborden la cantidad de datos disponibles, que consideren líneas espectrales adicionales para el entrenamiento y posterior predicción de las CNN, y que corrijan las condiciones de ruido evidenciadas en los espectros, podrían ser relevantes.

ANEXOS

5.1. Anexo A

Parameter	Set/Value
Convolutional Block	
kernel_size	{3, 5, 7, 9}
num_blocks	{1, 2, 3}
num_layers	{1, 2, 3, 4, 5}
filters	{32, 64, 128, 256, 512}
max_pooling	True
separable	False
Dense Block	
dense_layers	{1, 2, 3}
dense_units	{32, 64, 128, 256, 384, 512}
use_batchnorm	False
Regression Head	
output_dim	3
loss	'mse'
metrics	'mae'
AutoModel	
tuner	'bayesian'
max_trials	500

Tabla 16: Parámetros de entrenamiento Autokeras: *bayesian*.

5.2. Anexo B

Parameter	Set/Value
Convolutional Block	
kernel_size	{2, 3, 5}
num_blocks	{2, 3}
num_layers	{2, 3}
filters	{32, 64, 128, 256, 512}
max_pooling	True
separable	False
Dense Block	
dense_layers	{2, 3}
dense_units	{128, 256, 384, 512}
use_batchnorm	False
Regression Head	
output_dim	3
loss	'mse'
metrics	'mae'
AutoModel	
tuner	'greedy'
max_trials	300

Tabla 17: Parámetros de entrenamiento Autokeras: *greedy*.

5.3. Anexo C

HD	T_{eff}	T_{eff} Prediction	T_{eff} Error	$\log g$	$\log g$ Prediction	$\log g$ Error	$v \sin i$	$v \sin i$ Prediction	$v \sin i$ Error	χ^2
33328	19526	19316 ± 398	210 (1.1 %)	3.3	3.1 ± 0.0	0.2 (5.6 %)	287	302 ± 5	15 (5.1 %)	52
35165	17000	18675 ± 1332	1675 (9.9 %)	4.0	3.3 ± 0.2	0.7 (16.8 %)	240	250 ± 20	10 (4.0 %)	93
35411	25900	16527 ± 593	9373 (36.2 %)	4.5	4.6 ± 0.2	0.1 (2.1 %)	53	140 ± 12	87 (164.1 %)	138
35439	25300	25255 ± 2779	45 (0.2 %)	3.5	3.5 ± 0.2	0.0 (0.7 %)	260	240 ± 16	20 (7.6 %)	62
37041	27500	21284 ± 4919	6216 (22.6 %)	3.7	3.1 ± 0.3	0.6 (17.0 %)	140	158 ± 35	18 (12.9 %)	140
37795	12200	29825 ± 316	17625 (144.5 %)	3.5	3.0 ± 0.1	0.5 (13.9 %)	-	220 ± 10	-	55
41335	22500	15645 ± 584	6855 (30.5 %)	3.4	3.4 ± 0.3	0.0 (1.3 %)	330	348 ± 17	18 (5.4 %)	39
42167	11430	23769 ± 291	12339 (108.0 %)	3.0	2.5 ± 0.2	0.5 (18.2 %)	-	242 ± 10	-	17
45725	19600	18201 ± 533	1399 (7.1 %)	3.4	3.2 ± 0.1	0.2 (6.2 %)	280	316 ± 9	36 (12.9 %)	56
45910	21300	19583 ± 3443	1717 (8.1 %)	3.9	3.4 ± 0.4	0.6 (14.0 %)	100	119 ± 30	19 (19.1 %)	343
48917	22130	16838 ± 1334	5292 (23.9 %)	3.3	3.4 ± 0.2	0.1 (1.7 %)	200	174 ± 17	26 (13.2 %)	89
50013	26000	16310 ± 489	9690 (37.3 %)	3.9	4.2 ± 0.3	0.3 (8.6 %)	290	253 ± 22	37 (12.8 %)	79
52918	25450	17157 ± 630	8293 (32.6 %)	4.5	3.7 ± 0.3	0.8 (16.8 %)	242	279 ± 16	37 (15.2 %)	78
56014	19300	18164 ± 316	1136 (5.9 %)	3.8	3.6 ± 0.1	0.2 (4.7 %)	200	180 ± 10	20 (10.1 %)	166
56139	17170	16567 ± 1305	603 (3.5 %)	3.9	3.2 ± 0.1	0.7 (18.6 %)	100	106 ± 12	6 (5.9 %)	162
57150	22000	16783 ± 1362	5217 (23.7 %)	3.5	3.1 ± 0.2	0.4 (10.1 %)	180	194 ± 16	14 (7.7 %)	91
57219	19300	19899 ± 375	599 (3.1 %)	4.0	3.7 ± 0.0	0.3 (8.2 %)	50	90 ± 5	40 (80.6 %)	219
57682	30000	16192 ± 408	13808 (46.0 %)	4.0	4.0 ± 0.2	0.0 (0.9 %)	10	43 ± 13	33 (325.4 %)	233
58343	20000	17254 ± 336	2746 (13.7 %)	3.6	3.3 ± 0.1	0.3 (8.4 %)	10	62 ± 6	52 (522.4 %)	214
58715	11820	24987 ± 254	13167 (111.4 %)	3.2	2.4 ± 0.2	0.8 (23.8 %)	-	238 ± 7	-	12
60606	18000	16972 ± 677	1028 (5.7 %)	3.5	3.7 ± 0.2	0.2 (4.8 %)	250	233 ± 11	17 (6.9 %)	92
63462	26000	21538 ± 538	4462 (17.2 %)	2.9	2.4 ± 0.1	0.5 (16.7 %)	300	320 ± 12	20 (6.7 %)	14
68423	12100	23479 ± 1079	11379 (94.0 %)	4.0	2.9 ± 0.2	1.1 (28.3 %)	-	106 ± 15	-	103
68980	23668	17819 ± 332	5849 (24.7 %)	3.3	2.8 ± 0.0	0.5 (14.8 %)	110	139 ± 5	29 (26.7 %)	102
71510	17643	18234 ± 268	591 (3.4 %)	4.1	3.6 ± 0.1	0.5 (11.4 %)	150	156 ± 7	6 (3.7 %)	172
75311	23895	17714 ± 380	6181 (25.9 %)	4.0	3.4 ± 0.1	0.6 (14.6 %)	250	257 ± 8	7 (2.9 %)	92
78764	19000	17376 ± 1079	1624 (8.5 %)	3.8	3.6 ± 0.2	0.2 (4.2 %)	140	147 ± 15	7 (4.8 %)	159
83953	15000	19009 ± 2907	4009 (26.7 %)	2.5	2.6 ± 0.2	0.0 (1.8 %)	250	258 ± 12	8 (3.0 %)	33
89080	11600	14925 ± 1567	3325 (28.7 %)	3.5	1.7 ± 0.1	1.8 (51.0 %)	-	252 ± 13	-	12
89890	15000	15444 ± 453	444 (3.0 %)	3.0	3.1 ± 0.1	0.1 (2.3 %)	27	65 ± 8	38 (143.2 %)	144
91465	17410	16338 ± 564	1072 (6.2 %)	3.0	3.0 ± 0.1	0.0 (0.5 %)	280	274 ± 11	6 (2.1 %)	58
92938	19000	16828 ± 1156	2172 (11.4 %)	3.5	3.4 ± 0.2	0.1 (3.5 %)	110	129 ± 13	19 (16.9 %)	130
93563	15000	16929 ± 1831	1929 (12.9 %)	2.3	2.1 ± 0.1	0.2 (9.3 %)	280	254 ± 8	26 (9.4 %)	22
98058	8200	17294 ± 453	9094 (110.9 %)	3.0	1.6 ± 0.1	1.4 (46.2 %)	-	197 ± 8	-	34
102776	20000	18128 ± 576	1872 (9.4 %)	3.2	3.4 ± 0.1	0.2 (4.9 %)	200	227 ± 6	27 (13.7 %)	81
103192	10980	14482 ± 304	3502 (31.9 %)	2.5	1.7 ± 0.0	0.9 (34.4 %)	-	276 ± 6	-	8
105382	15000	14701 ± 291	299 (2.0 %)	3.4	3.5 ± 0.2	0.1 (3.3 %)	68	103 ± 10	35 (51.9 %)	67
105435	22150	17282 ± 793	4868 (22.0 %)	3.4	3.4 ± 0.1	0.0 (0.7 %)	250	246 ± 7	4 (1.8 %)	66
107348	12830	15484 ± 1740	2654 (20.7 %)	2.3	1.9 ± 0.1	0.4 (17.4 %)	-	277 ± 8	-	22
110335	10600	25476 ± 677	14876 (140.3 %)	2.9	2.6 ± 0.2	0.3 (10.3 %)	-	231 ± 11	-	36
110432	22000	22560 ± 1656	560 (2.5 %)	3.0	2.7 ± 0.1	0.3 (11.2 %)	400	353 ± 20	47 (11.7 %)	21
112078	16500	18426 ± 2011	1926 (11.7 %)	3.0	3.0 ± 0.1	0.0 (0.0 %)	290	287 ± 16	3 (1.1 %)	55
120324	23950	17389 ± 184	6561 (27.4 %)	4.0	3.8 ± 0.1	0.2 (5.9 %)	110	131 ± 5	21 (18.9 %)	163
121492 ¹	4736	-	-	2	-	-	-	-	-	-
124195	15000	21406 ± 4762	6406 (42.7 %)	4.0	2.9 ± 0.3	1.1 (27.3 %)	150	200 ± 28	50 (33.3 %)	92
124367	19650	16347 ± 450	3303 (16.8 %)	3.0	3.0 ± 0.1	0.0 (1.5 %)	260	260 ± 10	0 (0.1 %)	58
124771	17100	17727 ± 192	627 (3.7 %)	3.5	3.3 ± 0.0	0.2 (5.8 %)	150	200 ± 6	50 (33.6 %)	114
127972	21000	16471 ± 289	4529 (21.6 %)	4.0	3.9 ± 0.2	0.0 (0.8 %)	240	219 ± 8	21 (8.9 %)	79
131492	20000	21379 ± 851	1379 (6.9 %)	4.0	4.0 ± 0.1	0.0 (0.8 %)	100	168 ± 14	68 (68.4 %)	188
135734	13500	15149 ± 398	1649 (12.2 %)	3.8	1.8 ± 0.0	2.0 (52.6 %)	-	232 ± 5	-	18
138769	18400	17335 ± 332	1065 (5.8 %)	3.5	3.9 ± 0.1	0.4 (12.8 %)	30	86 ± 8	56 (187.9 %)	188
142184 ²	17375	-	-	4.3	-	-	300	-	-	-
142983	18000	27315 ± 395	9315 (51.7 %)	3.7	3.9 ± 0.1	0.2 (5.2 %)	370	338 ± 9	32 (8.6 %)	49
143275	30900	19959 ± 522	10941 (35.4 %)	3.5	2.5 ± 0.1	1.0 (29.0 %)	257	250 ± 13	7 (2.8 %)	25
148184	30000	21321 ± 2132	8679 (28.9 %)	3.5	2.9 ± 0.1	0.6 (18.2 %)	150	168 ± 11	18 (12.2 %)	78
157042	22000	24480 ± 487	2480 (11.3 %)	3.9	3.8 ± 0.1	0.1 (1.6 %)	280	292 ± 7	12 (4.4 %)	66
157246	21240	21646 ± 304	406 (1.9 %)	3.5	3.4 ± 0.0	0.1 (1.6 %)	230	239 ± 6	9 (4.0 %)	84
158427 ²	18110	-	-	3.8	-	-	285	-	-	-
167128	20000	19265 ± 660	735 (3.7 %)	3.8	3.1 ± 0.1	0.7 (17.2 %)	50	138 ± 10	88 (175.5 %)	178
205637	20000	20824 ± 404	824 (4.1 %)	4.1	3.8 ± 0.1	0.3 (8.1 %)	230	231 ± 9	1 (0.6 %)	142
209014	12200	16253 ± 564	4053 (33.2 %)	3.3	1.9 ± 0.1	1.4 (43.6 %)	-	348 ± 11	-	31
209409	15000	22411 ± 4760	7411 (49.4 %)	3.0	2.8 ± 0.3	0.2 (5.4 %)	350	271 ± 17	79 (22.7 %)	53
212076	18000	16191 ± 254	1809 (10.0 %)	4.0	4.0 ± 0.2	0.0 (0.7 %)	100	121 ± 7	21 (20.6 %)	132
212571	24500	24656 ± 1740	156 (0.6 %)	3.4	3.4 ± 0.1	0.0 (1.3 %)	215	228 ± 8	13 (6.0 %)	78
214748	11500	16308 ± 2011	4808 (41.8 %)	3.9	1.8 ± 0.1	2.1 (53.5 %)	-	219 ± 16	-	18
217891	16000	16255 ± 1567	255 (1.6 %)	3.0	2.8 ± 0.1	0.2 (7.8 %)	90	105 ± 13	15 (16.7 %)	85
219688	14200	15496 ± 2779	1296 (9.1 %)	3.9	2.6 ± 0.2	1.3 (32.8 %)	-	277 ± 16	-	45
221507	11000	13427 ± 404	2427 (22.1 %)	4.4	1.5 ± 0.1	2.9 (65.2 %)	-	229 ± 9	-	20

Tabla 18: Predicción del ensamble de modelos sobre espectros de estrellas Be.

¹El espectro de esta estrella no se encontraba dentro de la base de datos de BeSOS.

²El espectro de estas estrellas tiene un *step* de medición mayor al requerido para ser procesado por la red neuronal, por tanto se omitieron sus resultados.

5.4. Anexo D

Pos.	Layer Type	Dim 1 (/ Dim 2) / Filters	Parameters
1	InputLayer	(None, 139, 1)	0
2	CastToFloat32	(None, 139, 1)	0
3	Conv1D	(None, 138, 256)	768
4	Conv1D	(None, 137, 512)	262656
5	MaxPooling1D	(None, 137, 512)	0
6	Dropout	(None, 137, 512)	0
7	Conv1D	(None, 136, 32)	32800
8	Conv1D	(None, 135, 256)	16640
9	MaxPooling1D	(None, 135, 256)	0
0	Dropout	(None, 135, 256)	0
11	Flatten	(None, 34560)	0
12	Dense	(None, 384)	13271424
13	ReLU	(None, 384)	0
14	Dense	(None, 512)	197120
15	ReLU	(None, 512)	0
16	Dense	(None, 512)	262656
17	ReLU	(None, 512)	0
18	Dense	(None, 3)	1539
	Optimizer	Adam	

Tabla 19: CNN Optimizada para datos acotados con ruido. Tabla de especificaciones.

5.5. Anexo E

Parameter	Set/Value
Convolutional Block 1 & 2 ¹	
kernel_size	{2, 3, 5, 7}
num_blocks	{1}
num_layers	{1, 2}
filters	{128, 256}
max_pooling	True
separable	False
Dense Block	
dense_layers	{2, 3}
dense_units	{128, 256}
use_batchnorm	False
Regression Head	
output_dim	3
loss	'mse'
metrics	'mae'
AutoModel	
tuner	'bayesian'
max_trials	50

Tabla 20: Parámetros de entrenamiento Autokeras: verificación para datos de rango extendido y sin ruido.

5.6. Anexo F

Pos.	Layer Type	Dim 1 (/ Dim 2) / Filters	Parameters
1	InputLayer	(None, 701, 1)	0
2	CastToFloat32	(None, 701, 1)	0
3	Conv1D	(None, 699, 128)	512
4	MaxPooling1D	(None, 349, 128)	0
5	Conv1D	(None, 347, 128)	49280
6	MaxPooling1D	(None, 173, 128)	0
7	Dropout	(None, 173, 128)	0
8	Flatten	(None, 22144)	0
9	Dense	(None, 256)	5669120
10	ReLU	(None, 256)	0
12	Dense	(None, 256)	65792
13	ReLU	(None, 256)	0
14	Dense	(None, 128)	32896
15	ReLU	(None, 128)	0
16	Dense	(None, 3)	387
	Optimizer	Adam	

Tabla 21: CNN Optimizada para datos de rango extendido y sin ruido. Tabla de especificaciones.

¹En este caso se usaron dos bloques convolucionales con idénticos parámetros.

5.7. Anexo G

Parameter	Value
IMODE	0
IDSTD	42
IPRIN	3
Input Model	1
INTRPL	0
ICHANGE	1
ICHEMC	0
IOPHLI	0
IFREQ	0
INLTE	1
ICONTL	0
INLISR	0
IFHE2	1
IHYDPR	2122
IHE1PR	23
IHE2PR	24
ALAM0	4440.0
ALAST	4510.0
CUTOF0	91
CUTOFFS	1
RELOP	1.0D-4
SPACE	0.1
VTB	-99.
NMU0	0
ANG0	0
IFLUX	0

Tabla 22: Parámetros para generación de espectros de Synspec.

5.8. Anexo H

Parameter	Value
fname7	fort.7
fname17	fort.17
fnout	output
VROT	450
CHARD	0.01
STEPR	0.1
FWHM	0.000
STEPI	0
WSTART	4440.0
WEND	4510.0
IREL	1

Tabla 23: Parámetros para generación de espectros de Rotin.

REFERENCIAS BIBLIOGRÁFICAS

- [Abadi *et al.*, 2016] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., e Isard, M. (2016). Tensorflow: a system for large-scale machine learning. En *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pp. 265–283.
- [Allen, 2005] Allen, J. (2005). The classification of stellar spectra. http://www.star.ucl.ac.uk/~pac/spectral_classification.html. Acceso en Mayo 8, 2024.
- [Allende Prieto, 2004] Allende Prieto, C. (2004). Automated analysis of stellar spectra. *Astronomische Nachrichten: Astronomical Notes*, 325(6-8):604–609.
- [Arcos *et al.*, 2018] Arcos, C., Kanaan, S., Chávez, J., Vanzi, L., Araya, I., y Curé, M. (2018). Stellar parameters and $h\alpha$ line profile variability of be stars in the besos survey. *Monthly Notices of the Royal Astronomical Society*, 474(4):5287–5299.
- [Arrieta *et al.*, 2020] Arrieta, A. B., D’íaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garc’ía, S., Gil-L’opez, S., Molina, D., y Benjamins, R. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115.
- [Arvind *et al.*, 2023] Arvind, K., Besler, E., y Shi, L. (2023). 10 ensemble modeling.
- [Bailer-Jones, 2001] Bailer-Jones, C. A. (2001). Automated stellar classification for large surveys: a review of methods and results. *arXiv preprint astro-ph/0102223*.
- [Buduma *et al.*, 2022] Buduma, N., Buduma, N., y Papa, J. (2022). *Fundamentals of deep learning*. O’Reilly Media, Inc.”.
- [Cai *et al.*, 2023] Cai, B., Kong, X., Shi, J., Gao, Q., Bu, Y., y Yi, Z. (2023). Li-rich giants identified from lamost dr8 low-resolution survey. *The Astronomical Journal*, 165(2):52.
- [Decin *et al.*, 2004] Decin, L., Shkedy, Z., Molenberghs, G., Aerts, M., y Aerts, C. (2004). Estimating stellar parameters from spectra-i. goodness-of-fit parameters and lack-of-fit test. *Astronomy & Astrophysics*, 421(1):281–294.
- [Deisenroth *et al.*, 2020] Deisenroth, M. P., Faisal, A. A., y Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.
- [Der Kiureghian y Ditlevsen, 2009] Der Kiureghian, A. y Ditlevsen, O. (2009). Aleatory or epistemic? does it matter? *Structural safety*, 31(2):105–112.
- [Dodwell, 2023] Dodwell, T. (2023). Understanding uncertainty quantification: The different types.

- [Dorogush *et al.*, 2018] Dorogush, A. V., Ershov, V., y Gulin, A. (2018). Catboost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.
- [Fabbro *et al.*, 2018] Fabbro, S., Venn, K., O’Brian, T., Bialek, S., Kielty, C., Jahandar, F., y Monty, S. (2018). An application of deep learning in the analysis of stellar spectra. *Monthly Notices of the Royal Astronomical Society*, 475(3):2978–2993.
- [Folkes *et al.*, 1996] Folkes, S., Lahav, O., y Maddox, S. (1996). An artificial neural network approach to the classification of galaxy spectra. *Monthly Notices of the Royal Astronomical Society*, 283(2):651–665.
- [Garnett, 2023] Garnett, R. (2023). *Bayesian optimization*. Cambridge University Press.
- [Goodfellow *et al.*, 2017] Goodfellow, I., Bengio, Y., y Courville, A. (2017). *Deep learning*, volumen 1. MIT press Cambridge, MA, USA.
- [Gutiérrez Albores y Luna Aguilar, 2004] Gutiérrez Albores, L. y Luna Aguilar, E. (2004). Telescopios e instrumentación para la observación astronómica. <http://www.revista.unam.mx/vol.5/num4/art23/art23.htm>. Acceso en Mayo 7, 2024.
- [Hastie *et al.*, 2009] Hastie, T., Tibshirani, R., Friedman, J. H., y Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volumen 2. Springer.
- [Hubeny, 1988] Hubeny, I. (1988). A computer program for calculating non-lte model stellar atmospheres. *Computer Physics Communications*, 52(1):103–132.
- [Hubeny y Lanz, 2017] Hubeny, I. y Lanz, T. (2017). A brief introductory guide to tlusty and synspec. *arXiv preprint arXiv:1706.01859*.
- [Ji *et al.*, 2023] Ji, W., Liu, C., Deng, L., Zhang, B., Li, J., Tian, H., y Li, J. (2023). All spectral type lamost spectra library (atlas). *The Astrophysical Journal Supplement Series*, 265(2):61.
- [Jin *et al.*, 2019] Jin, H., Song, Q., y Hu, X. (2019). Auto-keras: An efficient neural architecture search system. En *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1946–1956.
- [Katz *et al.*, 1998] Katz, D., Soubiran, C., Cayrel, R., Adda, M., y Cautain, R. (1998). On-line determination of stellar atmospheric parameters T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$ from elodie echelle spectra. i-the method. *arXiv preprint astro-ph/9806232*.
- [Kaya, 2018] Kaya, I. (2018). 10 ensemble modeling.
- [Ke *et al.*, 2017] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., y Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.

- [Kingma y Ba, 2014] Kingma, D. P. y Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [Ksoll *et al.*, 2020] Ksoll, V. F., Ardizzone, L., Klessen, R., Koethe, U., Sabbi, E., Robberto, M., Gouliermis, D., Rother, C., Zeidler, P., y Gennaro, M. (2020). Stellar parameter determination from photometry using invertible neural networks. *Monthly Notices of the Royal Astronomical Society*, 499(4):5447–5485.
- [Lanz y Hubeny, 2007] Lanz, T. y Hubeny, I. (2007). A grid of nlte line-blanketed model atmospheres of early b-type stars. *The Astrophysical Journal Supplement Series*, 169(1):83.
- [LeCun *et al.*, 2015] LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.
- [Lee *et al.*, 2008] Lee, Y. S., Beers, T. C., Sivarani, T., Prieto, C. A., Koesterke, L., Wilhelm, R., Fiorentin, P. R., Bailer-Jones, C. A., Norris, J. E., y Rockosi, C. M. (2008). The segue stellar parameter pipeline. i. description and comparison of individual methods. *The Astronomical Journal*, 136(5):2022.
- [Leung y Bovy, 2019] Leung, H. W. y Bovy, J. (2019). Deep learning of multi-element abundances from high-resolution spectroscopic data. *Monthly Notices of the Royal Astronomical Society*, 483(3):3255–3277.
- [Levy *et al.*, 2017] Levy, P. R., Lugo, G. G., Pérez, C., y Pooley, D. (2017). Análisis espectral de la estrella post-agb hd 172481. *Ciencia e Ingeniería*, 38(3):239–247.
- [Li y Lin, 2023] Li, X. y Lin, B. (2023). Estimating stellar parameters from lamost low-resolution spectra. *Monthly Notices of the Royal Astronomical Society*, 521(4):6354–6367.
- [Li *et al.*, 2014] Li, X., Wu, Q. J., Luo, A., Zhao, Y., Lu, Y., Zuo, F., Yang, T., y Wang, Y. (2014). Sdss/segue spectral feature analysis for stellar atmospheric parameter estimation. *The Astrophysical Journal*, 790(2):105.
- [Li *et al.*, 2021] Li, Z., Liu, F., Yang, W., Peng, S., y Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*.
- [Liang *et al.*, 2022] Liang, J., Bu, Y., Tan, K., Pan, J., Yi, Z., Kong, X., y Fan, Z. (2022). Estimation of stellar atmospheric parameters with light gradient boosting machine algorithm and principal component analysis. *The Astronomical Journal*, 163(4):153.
- [Loshchilov y Hutter, 2017] Loshchilov, I. y Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- [MacRae, 2019] MacRae, M. (2019). Trends in spectroscopy: A snapshot of notable advances and applications. <https://www.spectroscopyonline.com/view/trends-spectroscopy-snapshot-notable-advances-and-applications>. Acceso en Mayo 15, 2024.

- [Majewski *et al.*, 2017] Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., Prieto, C. A., Barkhouser, R., Bizyaev, D., Blank, B., Brunner, S., Burton, A., y Carrera, R. (2017). The apache point observatory galactic evolution experiment (apogee). *The Astronomical Journal*, 154(3):94.
- [McCulloch y Pitts, 1943] McCulloch, W. S. y Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133.
- [Minh *et al.*, 2022] Minh, D., Wang, H. X., Li, Y. F., y Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, pp. 1–66.
- [Nath *et al.*, 2018] Nath, M. P., Pandey, P., Somu, K., y Amalraj, P. (2018). Artificial intelligence & machine learning: the emerging milestones in software development. *IJRSI*, 5(IX):36–44.
- [Neal *et al.*, 2018] Neal, B., Mittal, S., Baratin, A., Tania, V., Scicluna, M., Lacoste-Julien, S., y Mitliagkas, I. (2018). A modern take on the bias-variance tradeoff in neural networks. *arXiv preprint arXiv:1810.08591*.
- [Ness *et al.*, 2015] Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y., y Zasowski, G. (2015). The cannon: A data-driven approach to stellar label determination. *The Astrophysical Journal*, 808(1):16.
- [Pawlikowski y Chorowska, 2020] Pawlikowski, M. y Chorowska, A. (2020). Weighted ensemble of statistical models. *International Journal of Forecasting*, 36(1):93–97.
- [Royer *et al.*, 2024] Royer, P., Merle, T., Dsilva, K., Sekaran, S., Van Winckel, H., Frémat, Y., Van der Swaelmen, M., Gebruers, S., Tkachenko, A., y Laverick, M. (2024). Melchior—the mercator library of high resolution stellar spectroscopy. *Astronomy & Astrophysics*, 681:A107.
- [Shkedy *et al.*, 2007] Shkedy, Z., Decin, L., Molenberghs, G., y Aerts, C. (2007). Estimating stellar parameters from spectra using a hierarchical bayesian approach. *Monthly Notices of the Royal Astronomical Society*, 377(1):120–132.
- [Simón-Díaz, 2020] Simón-Díaz, S. (2020). A modern guide to quantitative spectroscopy of massive ob stars. *Reviews in Frontiers of Modern Astrophysics: From Space Debris to Cosmology*, pp. 155–187.
- [Singh *et al.*, 1998] Singh, H. P., Gulati, R. K., y Gupta, R. (1998). Stellar spectral classification using principal component analysis and artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, 295(2):312–318.
- [Templeton, 2009] Templeton, M. (2009). The september 2009 variable star of the season. https://www.aavso.org/vsots_pcyg. Acceso en Mayo 8, 2024.
- [Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

- [Ting *et al.*, 2019] Ting, Y.-S., Conroy, C., Rix, H.-W., y Cargile, P. (2019). The payne: self-consistent ab initio fitting of stellar spectra. *The Astrophysical Journal*, 879(2):69.
- [Vanzi *et al.*, 2012] Vanzi, L., Chacon, J., Helminiak, K., Baffico, M., Rivinius, T., Štefl, S., Baade, D., Avila, G., y Guirao, C. (2012). Pucheros: a cost-effective solution for high-resolution spectroscopy with small telescopes. *Monthly Notices of the Royal Astronomical Society*, 424(4):2770–2777.
- [Wu *et al.*, 2011] Wu, Y., Luo, A.-L., Li, H.-N., Shi, J.-R., Prugniel, P., Liang, Y.-C., Zhao, Y.-H., Zhang, J.-N., Bai, Z.-R., y Wei, P. (2011). Automatic determination of stellar atmospheric parameters and construction of stellar spectral templates of the guoshoujing telescope (lamost). *Research in Astronomy and Astrophysics*, 11(8):924.
- [Xiang *et al.*, 2021] Xiang, G., Chen, J., Qiu, B., y Lu, Y. (2021). Estimating stellar atmospheric parameters from the lamost dr6 spectra with scdd model. *Publications of the Astronomical Society of the Pacific*, 133(1020):024504.
- [Xiang *et al.*, 2015] Xiang, M., Liu, X., Yuan, H., Huang, Y., Huo, Z., Zhang, H., Chen, B., Zhang, H., Sun, N., y Wang, C. (2015). The lamost stellar parameter pipeline at peking university–lsp3. *Monthly Notices of the Royal Astronomical Society*, 448(1):822–854.
- [Young *et al.*, 2008] Young, H. D., Freedman, R. A., y Ford, A. L. (2008). *Sears and Ze-mansky’s university physics*, volumen 2. Pearson education.
- [Zhang *et al.*, 2019] Zhang, X., Zhao, G., Yang, C., Wang, Q., y Zuo, W. (2019). 938,720 giants from lamost i: Determination of stellar parameters and α , c, n abundances with deep learning. *Publications of the Astronomical Society of the Pacific*, 131(1003):094202.
- [Zhao *et al.*, 2022] Zhao, Z., Wei, J., y Jiang, B. (2022). Automated stellar spectra classification with ensemble convolutional neural network. *Advances in Astronomy*, 2022:1–7.