applied sciences

MDPI

Article

Application of Deep Learning to Spectroscopic Features of the Balmer-Alpha Line for Hydrogen Isotopic Ratio Determination in Tokamaks

Mohammed Koubiti * and Malo Kerebel

PIIM Laboratory, Aix-Marseille University/CNRS, CEDEX 20, 13013 Marseille, France
* Correspondence: mohammed.koubiti@univ-amu.fr

Abstract: We propose in this paper the use of artificial intelligence, especially deep learning algorithms, for the isotopic ratio determination for hydrogen–deuterium mixtures. Our approach is based on the Balmer-α line emitted by hydrogen and deuterium, but unlike the standard method, it does not consist of fitting the Hα/Dα line spectra. Instead, only some basic spectroscopic features such as the line peak-to-dip wavelength separation, peak-to-peak and dip-to-peak intensity ratios of the Zeeman–Doppler-broadened Hα/Dα line spectra are used by the regression algorithm for training. We demonstrate the proof-of-principle of our approach by applying deep learning from the open-access machine-learning platform TensorFlow to Hα/Dα line profiles, which we have synthetized with pre-determined parameters such as neutral temperatures, the magnetic field strength and the H/(H+D) isotopic ratio. The used regression algorithm allowed us to retrieve with a good accuracy the isotopic ratios used for the synthetized line profiles.

Keywords: plasma spectroscopy; isotopic ratio; machine learning; artificial intelligence; deep-learning; fusion plasmas; tokamaks; hydrogen isotopes; Zeeman–Doppler line profiles

Citation: Koubiti, M.; Kerebel, M. Application of Deep Learning to Spectroscopic Features of the Balmer-Alpha Line for Hydrogen Isotopic Ratio Determination in Tokamaks. *Appl. Sci.* **2022**, *12*, 9891. https://doi.org/10.3390/app12199891

Academic Editor: Emilio Martines

Received: 18 August 2022
Accepted: 29 September 2022
Published: 1 October 2022

1. Introduction

Nowadays magnetic fusion devices operate with either pure hydrogen, pure deuterium or with hydrogen–deuterium mixtures. However, magnetic fusion-based power plants will operate with D-T mixtures as fusion reaction rates are optimal when D^+ and T^+ ions fuse together as compared to fusion between D^+ ions. Up to now, only a few operations have been conducted or are being prepared with deuterium–tritium (D-T) mixtures like in JET during the 1997 DTE1 campaign and the DTE2 planned experimental campaign [1–3]. However, due to the radioactivity of tritium, its content in any device is limited for obvious safety reasons. Therefore, the content of tritium is strictly controlled. For this reason, an accurate inventory of the content of tritium is mandatory. Such an inventory cannot be made by simply subtracting the quantity exhausted by the pumping systems from the injected quantity because of the wall retention of all hydrogen isotopes [4]. In plasmas with D-T mixtures, one can quantify the content of tritium by the isotopic ratio, defined as:

$$IR_{\text{D-T}} = \frac{n_{\text{D}}}{(n_{\text{D}} + n_{\text{T}})}. \tag{1}$$

This is the ratio between the density of deuterium atoms n_{D} and the total density of deuterium plus tritium atoms $(n_{\text{D}} + n_{\text{T}})$. Another way to define the isotopic ratio is to use:

$$IR_{\text{T-D}} = \frac{n_T}{(n_{\text{D}} + n_{\text{T}})}. \tag{2}$$

In practice, as:

$$IR_{\text{D-T}} + IR_{\text{T-D}} = 1, \tag{3}$$

one can use either the first or the second definitions of the isotopic ratio, i.e., $IR_{\text{D-T}}$ or $IR_{\text{T-D}}$. In a similar manner to D-T mixtures, one defines the following isotopic ratios for hydrogen–deuterium (H-D) mixtures:

$$IR_{\text{H-D}} = n_{\text{H}}/(n_{\text{H}} + n_{\text{D}}) \tag{4}$$

$$IR_{\text{D-H}} = n_{\text{D}}/(n_{\text{H}} + n_{\text{D}}) \tag{5}$$

In Equations (4) and (5), n_{H} stands for the neutral hydrogen density. Note also that $IR_{\text{H-D}}$ and $IR_{\text{D-H}}$ satisfy the following condition:

$$IR_{\text{H-D}} + IR_{\text{D-H}} = 1 \tag{6}$$

It should also be noted that in some references, instead of Equations (4) and (5) concise notations are used for the isotopic ratios for H-D mixtures:

$$IR_{\text{H-D}} = \frac{H}{(H + D)} \tag{7}$$

$$IR_{\text{D-H}} = \frac{D}{(H + D)} \tag{8}$$

The same remark applies to D-T mixtures. To our knowledge, two methods are used for the evaluation of such isotopic ratios in magnetic fusion devices. The first one is based on the residual gas analysis and have been applied by Drenik et al. [5] to JET and ASDEX-Upgrade (AUG) devices. This technique has also been used for other devices like Tore-Supra by Klepper et al. [6]. However, it is the second method which is widely and commonly used for the evaluation of the hydrogen isotope contents. It is based on the spectroscopic analysis of the emitted Balmer-alpha line. In tokamaks, this line is one of the most intense emission lines of hydrogen isotopes (H, D and T). The intensities of the corresponding Hα, Dα and Tα lines depend on the concentration of each isotope. In tokamaks, the major emission of these lines comes from the divertor and scrape-off-layer (SOL) regions. In these regions, especially in the divertor where plasma–surface interactions take place, several neutral populations co-exist due to the different recycling mechanisms taking place as can be found in Kubo et al. [7], Koubiti et al. [8] and Hey et al. [9]. Under typical divertor conditions, i.e., electron densities n_e of the order of 10^{14} cm^{-3} and electron temperatures T_e of a few eV, the spectral profiles of the hydrogen isotope Balmer-α lines Hα, Dα and Tα are governed by Zeeman and Doppler effects, as Stark effects can be neglected for such low electron densities; see Koubiti and Sheeba [10]. Therefore, by fitting an experimental Balmer-α line emitted by hydrogen isotopes, it is possible to infer various parameters like the temperatures and fractions of the different neutral populations in the emissive zone, and more importantly the fraction of each hydrogen isotope and, hence, the isotopic ratio. Therefore, by applying this spectroscopic technique to spectra measured along various lines of sights, one can draw a cartography of the content of a given hydrogen isotope of the considered device. For instance, Neverov et al. [11,12] have applied this technique to JET data. In the previous references [7–9], even if the main objective was to determine the different neutral populations and characterize the dominant recycling mechanisms, the fractions of hydrogen isotopes were also determined. Usually, it is necessary to consider the co-existence of at least two neutral populations, each with its velocity distribution for each hydrogen isotope to fit the observed spectra. The presence of the magnetic field may lead to even more complications if its intensity is not sufficiently enough that the strong field approximation can be used. It is obvious that this technique relies on the fit of the whole line profiles of the considered line, i.e., the Balmer-α emitted by hydrogen isotopes. Besides, for its cost in terms of CPU time, this technique is convenient for post-operation interpretation of the measured spectra and cannot be used for real-time plasma controls. In a nuclear fusion reactor working with D-T mixtures, a real-time control of the tritium content must be necessary to ensure good performances and meet safety conditions. In

this context, a fast method allowing the determination of the isotopic ratio is necessary. In this paper, we propose to combine artificial intelligence with plasma spectroscopy to determine the isotopic ratio of hydrogen in tokamaks and stellarators for two objectives: promote plasma control through the monitoring of the hydrogen isotope content and predict future operations from available experimental data. Here, we will only present the proof of principle of the proposed method. In this method, instead of searching to fit the whole spectrum of the Balmer-α line, we used only some of its spectroscopic characteristics as input features in a deep-learning algorithm to determine the isotopic ratio for H-D mixtures. It is not the first time that machine-learning is introduced in plasma physics and plasma spectroscopy. Indeed, in 2021, Kajita et al. [13] used a multiple regression algorithm from the open source Sickit-Learn package machine-learning library [14] to analyze helium emission lines to determine the plasma electron density and temperature in the linear divertor simulator Magnum-PSI [15]. In [15], Kajita et al. used the standard neutral helium line intensity ratio technique for electron density and temperature determination. This technique requires the combination of spectroscopic intensity measurements of specific helium lines and a collisional–radiative modeling of the excited upper levels involved in the considered radiative transitions. The same regression technique was also used by Nishijima, Kajita and Tynan [16] to predict electron density and temperature from helium lines, this technique being validated using independent measurements using Langmuir probes. In [13], the assessment was realized thanks to independent measurements made with a Thomson Scattering system. In other physics fields, machine learning has been used for several years. For instance, in astrophysics, Gurung-López et al. [17] have published a work about an open-source code using deep learning for fitting profiles of the hydrogen Lyman-alpha line emitted in galaxies. One can also refer to the work of Cianciosa et al. [18] on the use of machine learning for the analysis of atomic spectral lines. It may be useful to mention some other references dealing with machine-learning in plasma spectroscopy such as MacBride et al. [19], Ren et al. [20] or Samuell et al. [21]; this short list is not exhaustive. The introduction of machine learning in plasma spectroscopy is not aimed to replace physical models but to help increase the speed of obtaining the same information as the physical models with a good accuracy to allow real-time control and monitoring. In addition, it is possible to make predictions for conditions of inaccessibility and in preparation for future experiments. This paper is organized as follows: in Section 2, we describe the physical model behind the determination of the hydrogen isotopic ratio, i.e., the theoretical computation of the line profiles of the Hα/Dα line for H-D mixtures, and the results can be easily extrapolated to Dα/Tα lines in D-T mixture experiments. In Section 3, we briefly present the concepts of machine learning by considering supervised machine learning and deep machine learning with a focus on neural networks. In Section 4, we apply machine learning to the Balmer-α emission line spectroscopy to demonstrate the proof-of-concept of our approach. This will be done through the generation of about two hundred thousands of synthetic Hα/Dα line profiles computed with given magnetic field strengths, different couples of neutral temperatures (cold and warm populations) and a fraction of neutral hydrogen atoms varying from 1 to 25% for deuterium plasmas with hydrogen. In Section 5, we will discuss our technique, its limitations and its potential applications such as predicting the tritium isotopic ratio in ITER from the analysis of available experiments carried out on the various fusion devices, including AUG and JET, whether operated with a mixture of hydrogen and deuterium or of deuterium and tritium. A conclusion will be drawn in the Section 6.

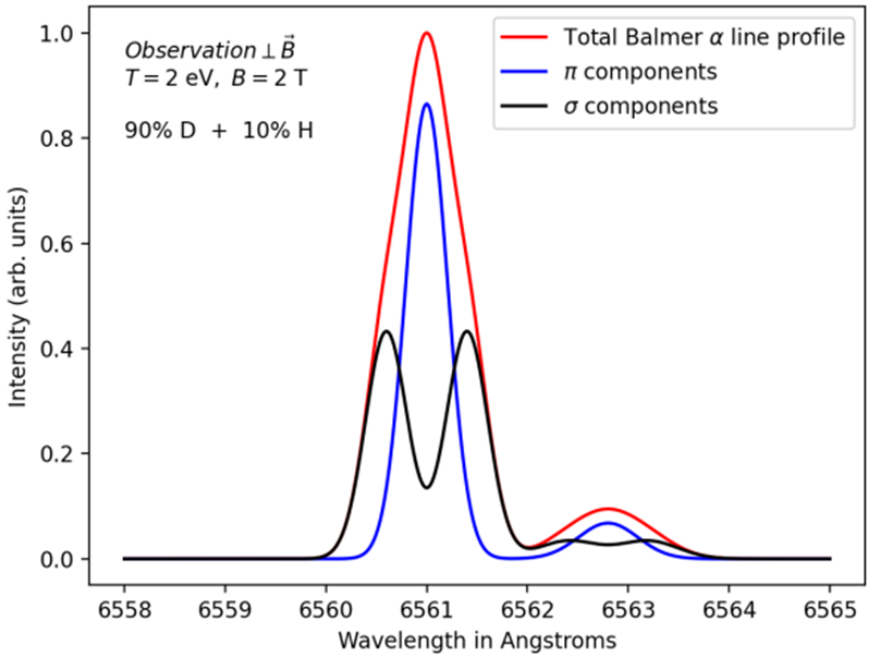
2. Classical Physical Model for the Hydrogen Isotopic Ratio Determination

For simplicity, we consider here tokamak plasmas where the dominant gas is deuterium with the presence of hydrogen as a trace up to a significant concentration of 25%. The results can easily be extended to the opposite situation, where the dominant gas is hydrogen with the presence of deuterium with concentrations varying between 1 to 25%. Therefore, we will consider here plasmas with H-D mixtures for which we aim to infer one

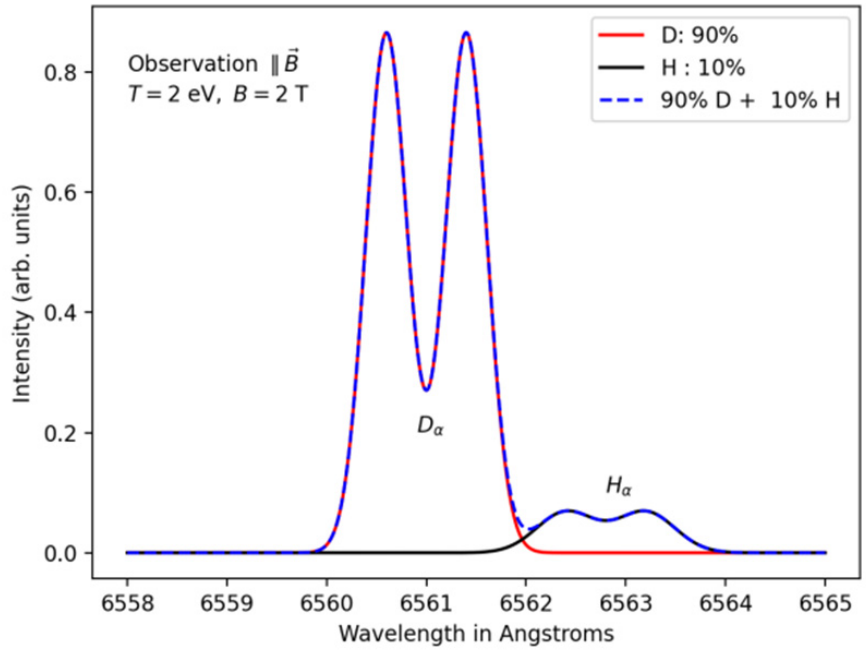
of the previously defined isotopic ratios: $IR_{\text{H-D}}$ or $IR_{\text{D-H}}$ (see Section 1). In such tokamak plasmas, the Balmer-α line is emitted by hydrogen and deuterium neutrals that are present in the tokamak divertor region. We will explain later why there exist neutrals and even molecules in the peripheral regions of tokamak. More precisely, Balmer-α photons are emitted as a result of electrons decaying from the upper excited level $n = 3$ of the atom to its lower excited level $n = 2$, n and n' being the principal quantum numbers. Because of the different reduced masses for hydrogen, deuterium and tritium, the wavelengths of the Hα/Dα/Tα emission lines are separated by a small isotope shift. In Angstrom units, the wavelengths of the Hα/Dα/Tα lines are respectively $\lambda_{H_\alpha} = 6562.8$, $\lambda_{D_\alpha} = 6561.04$ and $\lambda_{T_\alpha} = 6560.45$ [22,23]. This means that the non-perturbed wavelengths of Hα and Dα lines λ_{H_α} and λ_{D_α} are separated by about 1.8 Å while the Tα line is distant from the Dα line by only about 0.6 Å, i.e., $(\lambda_{D_\alpha} - \lambda_{T_\alpha})$ is about one-third of the separation $(\lambda_{H_\alpha} - \lambda_{D_\alpha})$ between Hα and Dα lines. Now, let us explain the origin of presence of hydrogen isotope neutrals in the peripheral regions of tokamaks. The interesting plasma of a tokamak is the high-temperature magnetically confined plasma called the core plasma. The confined plasma core does not contain any neutrals but only ions and electrons. The core plasma is confined inside closed magnetic flux surfaces. Besides these closed flux surfaces, there are open magnetic field lines which hit the plasma facing components. The charged particles rotating around these open field lines interact with the plasma wall materials, especially in the divertor region. The plasma–wall or plasma–surface interactions are at the origin of the presence of neutrals. More precisely, hydrogen and deuterium neutrals are produced through several chemical and physical sputtering mechanisms [7–9] leading to the co-existence of at least two neutral populations. The first is a cold neutral population resulting from molecular dissociation of desorbed molecules from the divertor target (release of trapped neutrals as molecules) whose velocity distribution can be considered as Maxwellian with a temperature of a fraction of eV to few eVs. The second neutral population originates from reflected atoms, i.e., H^+ or D^+ ions capture electrons when hitting the target, and are released (return back to the plasma) as H or D neutrals. Such a population can thermalize before emitting radiation (Hα/Dα photons) in which case their velocity distribution can be assumed to be Maxwellian (or close to a Maxwellian) with a temperature of some tens of eVs. A third neutral population due to charge exchanges between plasma ions and released neutrals may exist; its velocity distribution is also Maxwellian with a temperature of about 100 eV. This third population gives rise to a broad and weak line profile and therefore was not considered in our calculations of synthetic Hα/Dα line profiles.

In addition to the co-existence of the neutral populations described above for each hydrogen isotope, the profile of the emitted Balmer-α lines are also subject to the Zeeman effect caused by the magnetic field. It is out of the scope of this paper to describe in detail the effect of an external magnetic field on an emission line, as one can refer to standard textbooks treating this topic. Instead, we refer the readers to our previous paper [10] and the references therein for a brief description. However, it is widely accepted that for magnetic fields with strengths higher than about 1 T, one can use the strong magnetic field approximation, as the energy-level Zeeman splitting is much larger than the fine structure level splitting. Therefore, in the following, we use the strong magnetic field approximation in the computation of the perturbed energies of the hydrogen isotopes and Zeeman–Doppler broadening when calculating theoretical or synthetic H/D Balmer-α line profiles under different conditions, i.e., population fractions and isotopic rations, mainly. As the energy levels involved in the Balmer Hα/Dα lines have low principal quantum numbers and the electron densities we deal with in tokamaks are relatively low, we will simply ignore the Stark broadening, which is due to the electric micro-field created by the plasma-charged particles. Another crucial point is that the presence of an external magnetic field introduces anisotropy for the emitted photons. In fact, the isotropy of spontaneously emitted photons is no longer valid when considering spectral profiles of lines emitted in tokamaks, especially the divertor region. The shape of the Hα/Dα line depends strongly on the angle θ between the direction of the magnetic field and the observation direction, i.e.,

the line-of-sight. It is interesting to consider spectra at two specific values of the angle θ: 0 (or 180°) and 90°, which correspond respectively to parallel and perpendicular observations. The shape of the line profile for any other angle θ in the range $]0 : \frac{\pi}{2}[\cup]\frac{\pi}{2} : \pi[$ is somehow between the parallel and perpendicular profiles. For illustration, we show in Figure 1 the normalized profiles of the Hα/Dα lines calculated for a magnetic field of 2 T by considering only one neutral population having a Maxwellian VDF with a temperature of 2 eV. A gas mixture composed of 10% of hydrogen and 90% of deuterium was assumed for this illustrative calculation. We have used the strong magnetic field approximation for the Zeeman effect for this relatively high magnetic field strength. In addition, as the electron density n_e is typically of the order of 10^{14} cm^{-3} in tokamak divertors, we have neglected the Stark line, broadening in our calculations. In Figure 1a, the line profiles were calculated for a perpendicular observation ($\theta = 90°$) with respect to the magnetic field direction. One can see that for each hydrogen isotope, the total line profile represented by the red solid line is a sum of a central π component (represented by the blue solid line) located respectively at $\lambda_{D_\alpha} = 6561.04$ Å for Dα and $\lambda_{H_\alpha} = 6562.8$ for Hα lines and two lateral σ^+ and σ^- components (black solid line). The σ^+ and σ^- components are equally spaced from the central π component and their intensities are half of that of the π component. In Figure 1b, the shown profiles were calculated for a parallel observation with respect to the magnetic field direction, i.e., for an angle $\theta = 0°$ or 180°. In this case, the line profile of each hydrogen isotope consists onto the two lateral components σ^+ and σ^- and line dips are visible at the unperturbed wavelengths of the Balmer-α lines, i.e., at $\lambda_{D_\alpha} = 6561.04$ and $\lambda_{H_\alpha} = 6562.8$. To be a bit more comprehensive, we show in Figure 2 the synthetic Hα/Dα line profile calculated for a parallel observation by considering the same conditions as in Figure 1 but assuming a neutral population having a Maxwellian VDF with a neutral temperature of 20 eV. Note that each σ component is Doppler-broadened with a full width at half maximum FWHM more than three times (more precisely a factor of $\sqrt{10}$) than that of the previous case of Figure 1. For this temperature of 20 eV, the FWHM (~0.92 for Dα; ~0.65 for Hα) is greater than the $\pi - \sigma$ separation (~0.4 for both Dα and Hα). This explains why one observes single broad peaks centered at the unperturbed line wavelengths 6561.04 and 6562.8 . Note also that the $\pi - \sigma$ separation increases linearly with the strength of the magnetic field (~0.2 /T for each component σ^+ and σ^-). To close this section, we present in Figure 3 a more realistic calculation corresponding to the combination of Figures 1b and 2, but assuming the emission results from two neutral populations: a cold neutral population and a warm neutral one, both having a Maxwellian VDF but with different temperatures: 2 eV and 20 eV respectively. For this illustration, it was assumed that the cold population represents 75% while the warm population represents only 25%. This is true for hydrogen neutrals as well as deuterium neutrals. It should be noted here that because deuterium neutrals have a mass twice that of the hydrogen neutrals, the FWHM of the Hα line is greater by a factor of $\sqrt{2}$ than the FWHM of the Dα line. This is why the σ^+/σ^- peaks of the Dα line are well separated in comparison with those of the Hα line.



(**a**)



(**b**)

Figure 1. Zeeman–Doppler profiles of the Balmer-α lines (Hα/Dα) computed for a gas mixture composed of 10% of hydrogen and 90% of deuterium with a temperature of 2 eV in the presence of a magnetic field of 2 T. (**a**) Perpendicular observation ($\theta = 90°$): for each hydrogen isotope, the total line profile (red solid line) is a sum of a central π component (blue solid line) located respectively at $\lambda_{D_\alpha} = 6561.04$ Å for Dα and $\lambda_{H_\alpha} = 6562.8$ for Hα lines and two lateral σ^+ and σ^- components (black solid line). (**b**) Parallel observation: the spectrum was calculated for a parallel observation with respect to the magnetic field direction, i.e., for an angle $\theta = 0°$ or 180°.

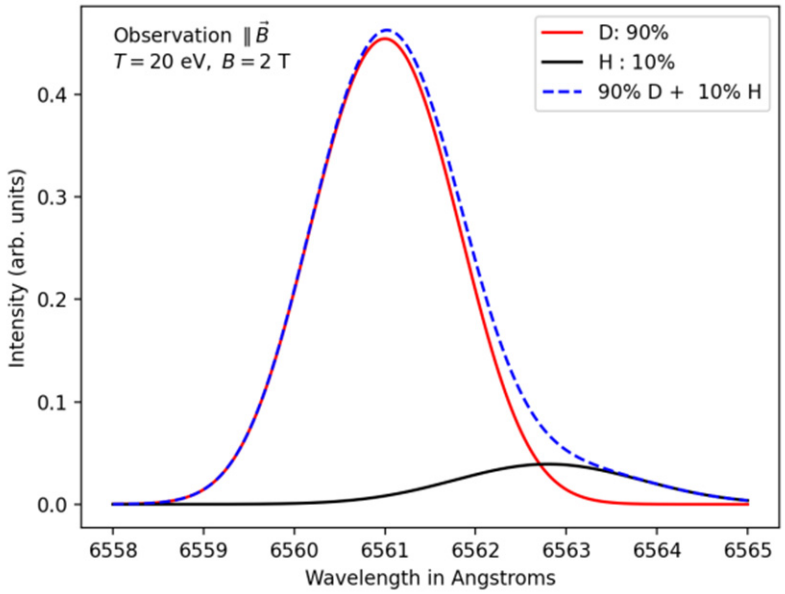


Figure 2. Same as Figure 1b (synthetic Hα/Dα line profile), being computed for a parallel observation with respect to the magnetic field, but for a neutral population having a Maxwellian VDF with a temperature of 20 eV instead of 2 eV.

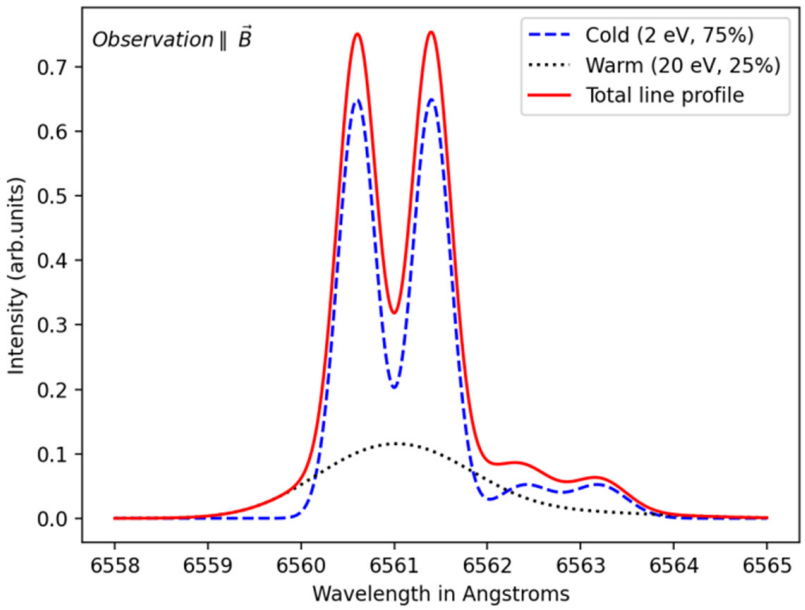


Figure 3. A synthetic Dα/Hα line profile calculated for a parallel observation with respect to the magnetic field direction whose intensity is 2 T. A 90%D + 10%H gas mixture was assumed for this calculation. The Balmer-α line emission profile (red solid line) is assumed resulting from the combination of two neutral populations having Maxwellian VDFs with temperatures of 2 eV (blue dashed line) and 20 eV (black dotted line). The relative fractions or concentrations of each population are as follows: 75% of cold neutrals and 25% of warm neutrals for each hydrogen isotope.

The method termed here as the classical method is based on the fit of the experimental Hα/Dα line spectra by modeling their theoretical profiles as explained previously in this section. In the following, to illustrate this physical model, we consider a synthetic profile instead of an experimental spectrum. The advantage of using a synthetic line profile is that there is no instrumental function to account for and as all the parameters are known, it is easy to assess the optimization technique allowing to fit the synthetic line profile. An example is illustrated in Figure 4. Synthetic data (synthetic spectrum calculated with

known parameters) are given as input to a lest-square minimization Python algorithm to be fitted, giving only the variation domains of the parameters (temperatures, fractions, isotopic ratio, magnetic field strength). As it can be seen, there is a very good agreement between the synthetic spectrum (full blue circles) and the fitted profile (red solid line). To demonstrate the robustness of this fitting algorithm on such synthetic spectra, we compared the parameters of the best fit to the known ones, i.e., the ones used to compute the synthetic Hα/Dα line spectrum. The known and deduced parameters are compared in Table 1. As it can be seen from Table 1, the inferred parameters are very close to the real ones, i.e., known ones. Of course, when it comes to the fit of real experimental spectra, things are more complex and other considerations should be taken into account. In particular, experimental line spectra contain a background, are generally noisy and lines are also subject to an additional broadening due to the instrumental function of the measurement system. However, in most situations, the instrumental function is Gaussian and therefore the spectra can be deconvoluted prior to their fitting, or an instrumental broadening should be considered in the fitting procedure. Fitting experimental Hα/Dα spectra from tokamak divertors is a widely used technique. In the last case, the obtained temperatures are "apparent temperatures". The real temperatures are then obtained from the apparent ones by subtracting a temperature value corresponding to the Gaussian broadening due to the instrumental function. In the next sections, instead of fitting such spectra, we propose to use only some spectroscopic features in an artificial intelligence algorithm to determine, in a fast way, the isotopic ratios for H-D mixtures in order to extrapolate them to D-T mixtures, make fast calculations for real-time operational control and make predictions for larger devices such as ITER and future fusion reactors.

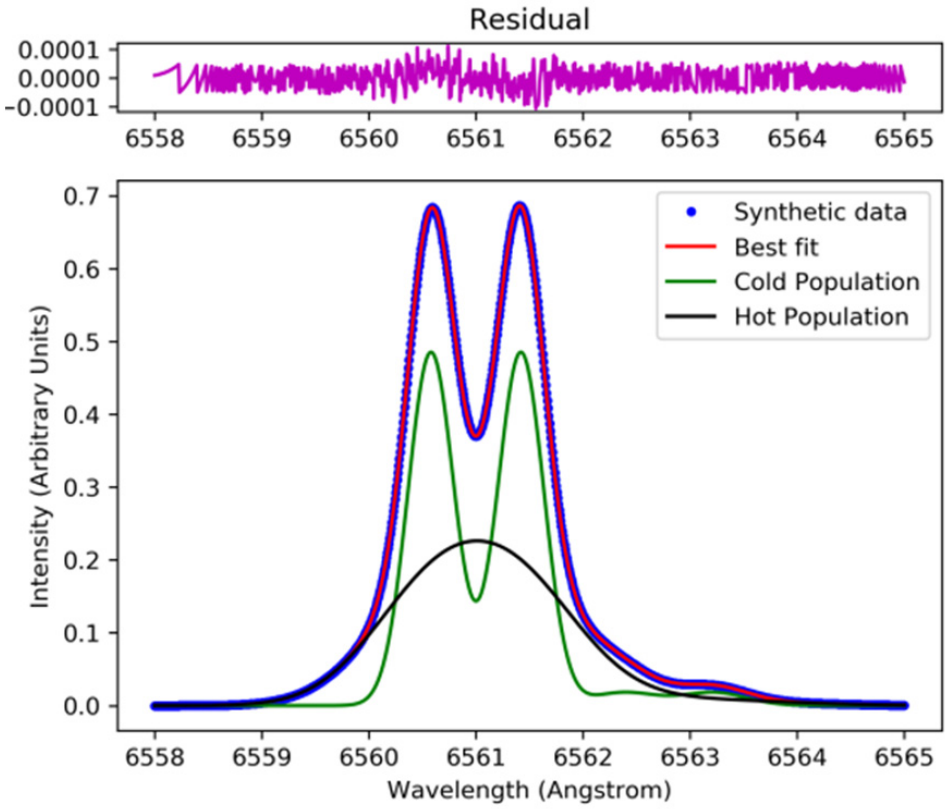


Figure 4. An example of the fitting of a synthetic Dα/Hα line profile calculated for a parallel observation with respect to the magnetic field direction. A least-square minimization Python algorithm was used to infer various parameters. The known and deduced parameters are summarized in Table 1.

Table 1. Comparison of the best fit parameters of the Python least-square minimization (fitting) algorithm with the real parameters (known parameters) of the synthetic line spectrum. T_C and T_W stand for the temperatures of the cold and warm neutral populations, respectively. f_C and f_W stand for their relative fractions in percentages.

Parameters	Known Values	Inferred Values	Errors
(T_C, f_C)	2 eV; 55%	2.015 eV; 54.998%	0.744%; 0.0042%
(T_W, f_W)	15 eV; 45%	15.111 eV; 45.002%	0.742%; 0.0051%
B	2 T	2.091 T	4.56%
$\frac{n_H}{n_H+n_D}$	5%	5.0007%	0.01308%
$\frac{n_D}{n_H+n_D}$	95%	94.9993%	0.00068%

3. Some Notions on Machine Learning and Deep Learning

In this section, we introduce, briefly, some basics of machine learning. Machine learning consists of the use of algorithms for classification, clustering, anomaly detection and regression issues. Classification consists of the identification of the category an object belongs to. Clustering consists of the automatic grouping of similar objects into sets. Algorithms of anomaly detection search to identify rare characteristics of a data sample element with respect to the majority of the data. Regression consists of predicting numerical values as a continuous function for an attribute associated with an object. There exist several algorithms such as random forest, nearest-neighbors or SVM (Support Vector Machines) for classification, or SVR (Support Vector Regression), nearest-neighbors and random forest for regression. For plasma physics applications, we use the regression feature of machine learning as we are interested by the prediction of physical quantities. For the results presented here, we have used the open-access machine-learning platform called TensorFlow [24] which is based on neural networks. We have precisely used the Python deep-learning API (Application Programming Interface) called Keras (see Keras website [25]). The regression algorithm which we have used is based on the ADAM optimizer (see Keras optimizers [26]). Note that there exist other machine learning platforms such as the open-access, Python-based Sickit-learn platform [27].

4. Application of Machine Learning to Spectroscopic Features of Hα/Dα Line Profiles

Machine learning algorithms require features and target data. For our application, we have used six input features. These are the strength B of the magnetic field $\vec{B}$, the temperatures T_C and T_W of the cold and warm neutral populations, the difference $\Delta\lambda_{H_a-\sigma^-(D)} = \lambda_{H_\alpha} - \lambda_{\sigma^-(D)}$ between the non-perturbed wavelength λ_{H_α} of the Hα line and the blue-shifted $\sigma^-(D)$ component of the Dα line, the ratio of the intensity of the Dα dip (or unshifted Dα wavelength λ_{D_α}) to that of its $\sigma^-(D)$ component, i.e., $\frac{I_{\mathrm{Dip}}(D)}{\mathrm{I}_{\sigma^-}(D)}$ and the ratio, $\frac{I_{\sigma^+(H)}}{\mathrm{I}_{\sigma^-}(D)}$ of the intensity of the $\sigma^+(H)$ component of the Hα line to of the $\sigma^-(D)$ component of the Dα line. The last three features are illustrated in Figure 5. The target data is the fraction of hydrogen neutrals or the isotopic ratio H/H+D. More precisely, we used a preprocessing program which computes many synthetic profiles by varying the magnetic field strength and the temperatures and relative fractions of both cold and warm H and D neutrals for a mixture of H and D with up to 25% of hydrogen. For the results presented here, we have used the typical following parameters: B in the range 1–5 T, cold and warm temperatures with the following respective values $T_C = 2 \pm 0.2$ eV and $T_W = 20 \pm 2$ eV, cold neutral fraction in the range 20–80% and a hydrogen concentration (isotopic ratio) in the range 1–25%. After calculating the line profiles, the preprocessing program extracts for each calculated profile the features described previously and transmits them to the regression algorithm. The whole process is illustrated in Figure 6.

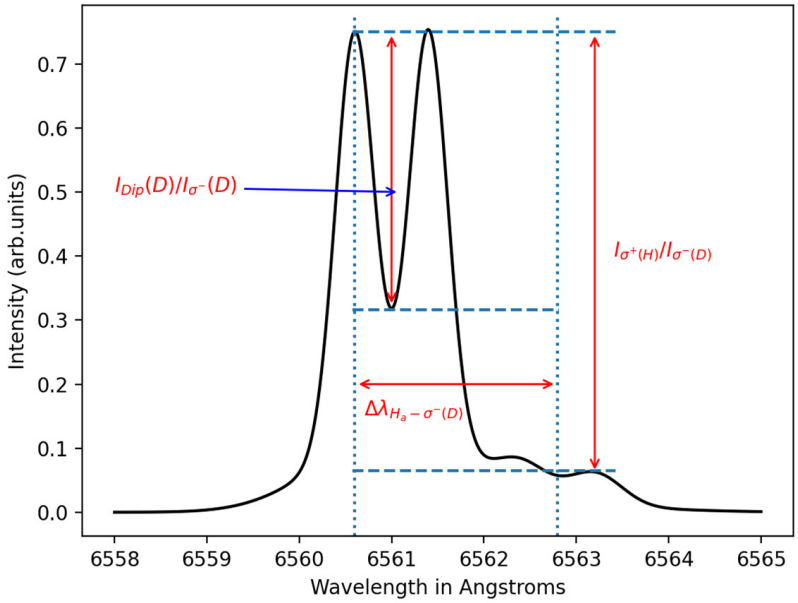


Figure 5. A scheme of a two-temperature Zeeman–Doppler Hα/Dα line profile indicating the spectroscopic characteristics which are part of the features data for the regression algorithm: the wavelength separation $\Delta\lambda_{H_\alpha-\sigma^-(D)}$ between the non-perturbed Hα wavelength $\lambda_{H_\alpha} = 6562.8$ and the blue-shifted (left component) σ^- component of the Dα line, the ratio of the intensity of the Dα dip to the intensity of its σ^- component, i.e., $\frac{I_{\mathrm{Dip}}(D)}{I_{\sigma^-}(D)}$, and the ratio of the intensity of the σ^+ component of the Hα line to of the σ^- component of the Dα line.

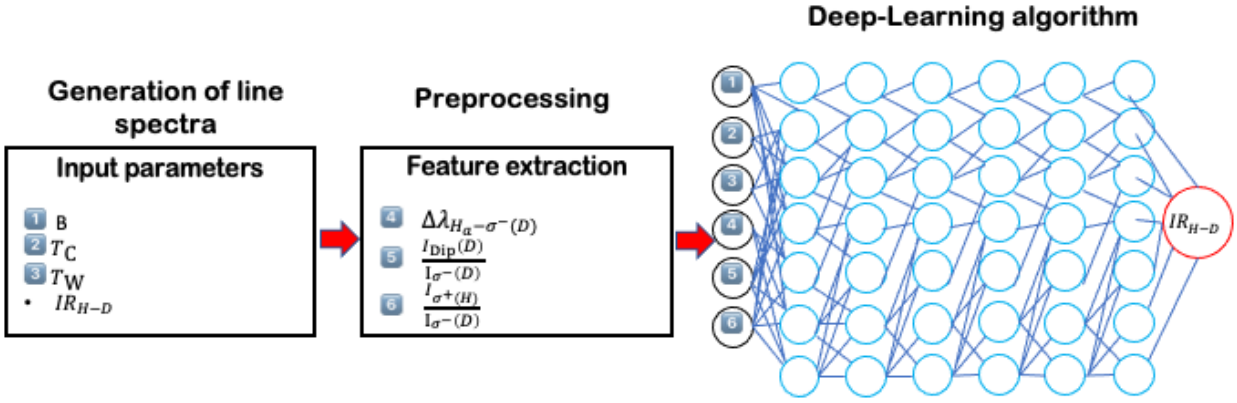


Figure 6. A scheme illustrating the whole method of the hydrogen isotopic ratio determination from the generation of theoretical Hα/Dα line profiles to the prediction of H/H+D ratios by deep learning (Neural Network algorithm). Numbers 1 to 6 inside the squares refer to the six input features.

The aim of the regression algorithm used here was to predict the isotopic ratio in a mixture of $(100 - x)\%D - x\%H$ where x varies between 1% and 25% by training it using the features mentioned above, in particular, the wavelength separation and the line intensity ratios, with the number of calculated spectra representing the number of samples. A set of 200,000 theoretical spectra were generated for the training phase of the process. Concerning the Neural network regression algorithm itself, it should be noted that in addition to the input and output neuron layers, our neural network is composed of six intermediate or hidden layers having respectively 150, 200, 400, 400, 200 and 100 neurons. The number of nodes or neurons of the input layer is equal to the number of input features, i.e., six in our case. The output layer contains a single node or neuron, as the target is a single numerical quantity (the isotopic ratio).

5. Results and Discussion

To illustrate the results of the application of the regression algorithm which we have adopted from the TensorFlow ML platform to our synthetized Hα/Dα line spectra, we show in Figure 7 the isotopic ratio (in percentages) as predicted by the ML algorithm versus the real isotopic ratio as used in the preprocessing code to generate the Hα/Dα line profiles. The loss function of the DL algorithm for our application is shown on Figure 8 as a function of the number of iterations or epochs. To obtain results corresponding to more realistic situations, we have added Gaussian noise to the synthetized Hα/Dα line profiles. The deep learning algorithm values are scattered around the black solid line representing the $y = x$ function. We found a mean error of 7.6% and a median error of 2.4% with respect to the expected values of the hydrogen isotopic ratio, H/H+D. Note that adding noise to the profiles makes the determination of the position of the peak of the Hα line difficult, and hence introduces an error in the ML feature wavelength separation. The consequence of this is that the regression algorithm introduces in turn an uncertainty on the isotopic ratios as can be seen on Figure 7. If one excludes outlier data, it can be reasonably considered that there is an overall good agreement between the predicted values of the isotopic ratio and the real values called true values in Figure 7. It is worth noting that the reasons of such relatively high error values are not due to the ML regression algorithm but are due to the difficulty of the preprocessing program in the determination of some features especially when the spectrum minima cannot be defined accurately. As this point sets limitations to the use of the regression algorithm and affects the accuracy of the predictions, it will be tackled in the near future to improve the proposed technique.

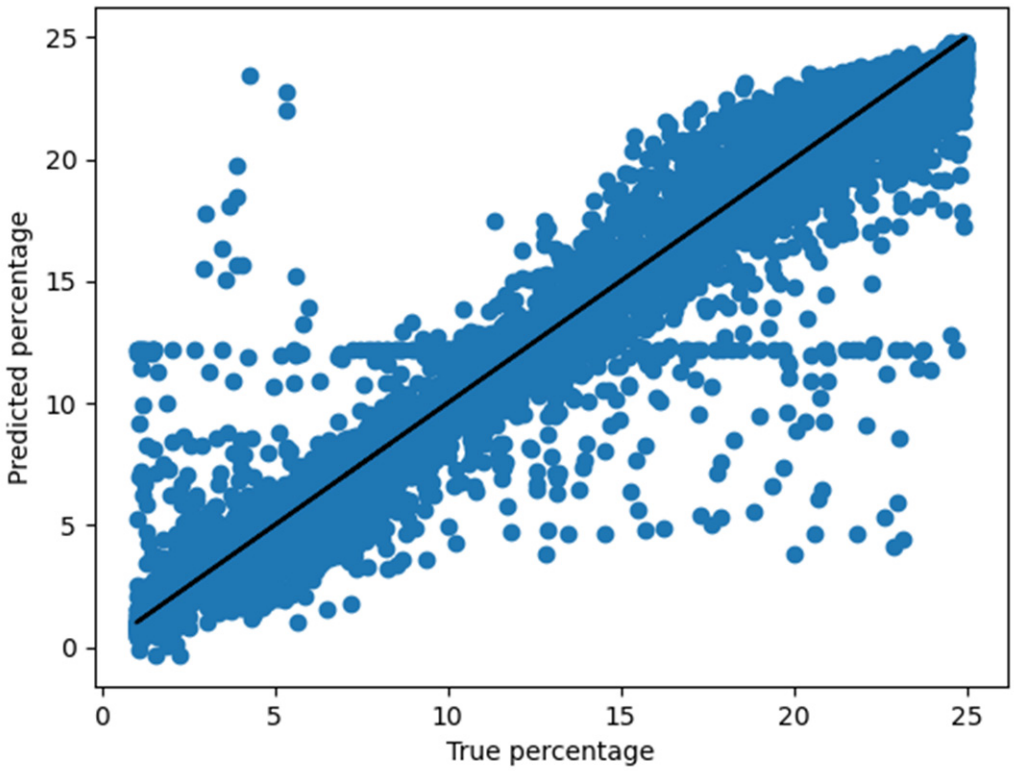


Figure 7. The predicted isotopic ratio as a function of the real (theoretical isotopic ratio) and the training set; a sample of 200,000 data points were used (200,000 theoretical line profiles were generated).

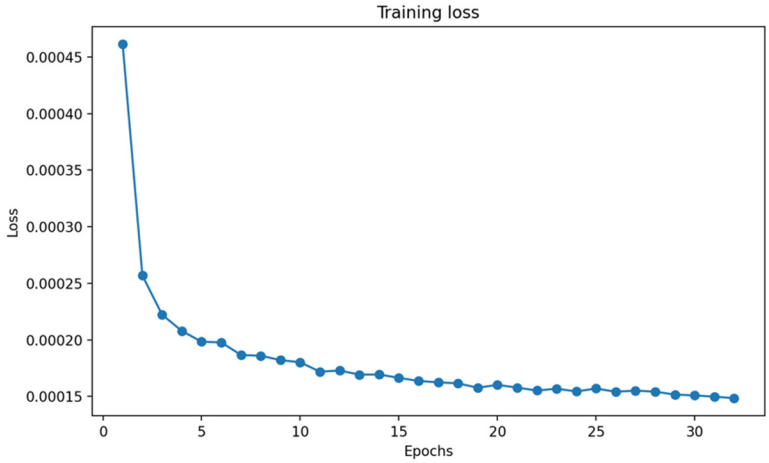


Figure 8. The loss function for the training data used for the determination of the hydrogen isotopic ratio. The less the loss is the better is the regression.

Now, concerning the validation of the method, we have generated a total of 20,000 other theoretical spectra in a similar mater as for the generation of the training set. The results of the DL algorithm are shown on Figure 9, as the predicted values of the hydrogen isotopic ratio are plotted as a function of the true values. For the test phase, the mean square error was estimated to be 7.9% while the median error value is 2.6%. One can see from the figure some outlier points for which the predicted values are far from the known values, but for the majority of the data, the predicted values match well the known values.

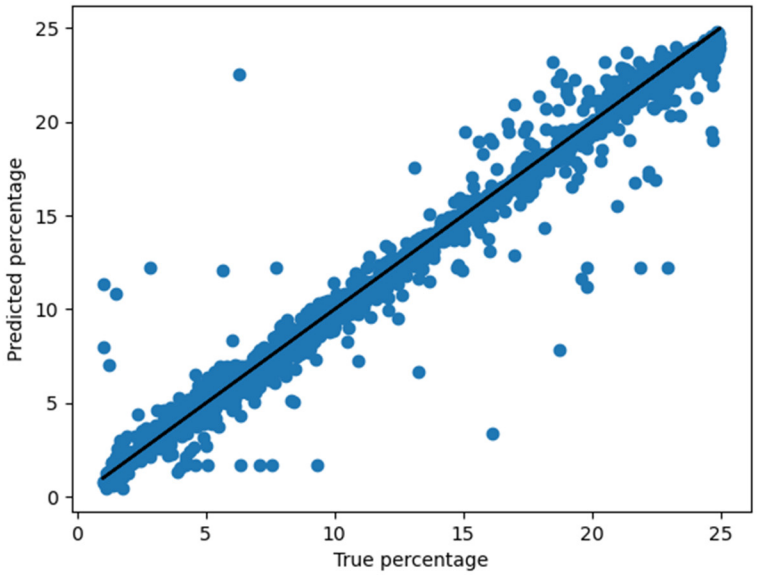


Figure 9. The predicted isotopic ratio as a function of the real (theoretical isotopic ratio) for the test set where a sample of 20,000 data points were used (20,000 theoretical line profiles).

The results presented here demonstrate that one can use simple spectral features of line spectra such as wavelength separation and dip-to-peak or peak-to-peak intensity ratios instead of fitting whole spectra to determine and predict isotopic ratios in H-D mixtures for conditions of divertor plasmas. We have limited our study to deuterium–hydrogen mixtures with hydrogen concentrations lying in the range 1–25%, but the results can be easily verified for the opposite situation where deuterium represents up to 25%, and also mixtures where hydrogen represents up to 50%. It should be noted that our results aimed to demonstrate the proof-of-principle of the use of ML coupled to spectroscopy to determine the isotopic ratio in H-D mixtures met in divertor plasmas of magnetic fusion devices such

as tokamaks. The next step will be the application of the method to experimental data from current tokamaks such as JET and ASDEX-Upgrade. In this step, we will limit ourselves to spectra measured in a direction parallel or quasi-parallel to the magnetic field direction; spectra measured perpendicularly to the magnetic field do not present separate σ^+/σ^- components because of the presence of the adjacent π component of both Hα and Dα lines. Furthermore, the addition of the plasma pressure, plasma temperature and density as well as other quantities as input features for the ML regression algorithm may allow us to discover some hidden connections with the isotopic ratio; this can be easily seen in the correlations between the different features and the target data (isotopic ration) in this case. Another point consists in the extrapolation of the results to D-T mixtures, as future DEMO and fusion nuclear power plants (reactors) will be operated with a mixture of deuterium and tritium. Determining the isotopic ratio D/D+T in fast way will be an important advantage for the device operators to control the plasma contents to be in line with the recommendation of the safety authorities and manage the performances of the reactors. All these issues not tackled in this paper will be the subject of another paper in the near future.

6. Conclusions

In this paper, we have demonstrated the proof-of-principle of a new technique to determine the isotopic ratios of hydrogen isotopes of interest for magnetic fusion research. In this technique, we combined simple spectral characteristics of the Balmer-α line emitted by hydrogen–deuterium gas mixtures with a deep learning regression algorithm adopted from the open-access machine-learning platform TensorFlow with Keras as the Python API. The demonstration was based on results obtained with Hα/Dα line profiles which we have computed varying a number of parameters such as the strength of the magnetic-filed B, the temperatures and fractions of the cold and warm populations of neutrals, reflecting the recycling mechanisms in tokamak divertors, and the concentration of hydrogen (in the range 1–25%). These parameters were varied to represent typical conditions of tokamak divertors. Instead of fitting whole spectra, only three parameters of each synthetic profile were used and transmitted to the regression algorithm for training: a wavelength separation between the unshifted wavelength of the Hα line and the blue-shifted σ^- of the Dα line, and dip-to-peak and peak-to-peak line intensity ratios. It has been shown that when the analyzed profiles don't contain noise, the regression algorithm predicts very accurately the expected isotopic ratios. However, when noise was introduced in the synthetized profiles, the predicted values were less accurate. The dispersion of the predictions around the expected values is not due to the deep learning treatment of the features provided to it but is attributed to the uncertainty in the determination of the spectral minima allowing to calculate the ML features. In other words, the error in the determination of the feature concerning the wavelength separation impacts the predictions of the deep learning algorithm. The next step of this work will be its application to available experimental data from tokamaks operating with H-D mixtures, and later for D-T plasmas.

Author Contributions: M.K. (Mohammed Koubiti) developed the code to synthetize Zeeman-Doppler broadened Ha/Da line profiles; M.K. (Malo Kerebel) developed the preprocessing program to extract the spectral features and adapted the deep-learning regression algorithm from TensorFlow-Keras to predict hydrogen isotopic ratios. M.K. (Mohammed Koubiti) supervised the work, wrote the manuscript. M.K. (Malo Kerebel), reviewed the manuscript before its submission. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the A*Midex IDEX, PIA AMX-19-IET-013 (Master's internship of M. Kerebel) as well the French Federation of Research on Magnetic Fusion (FR-FCM), Project 3IPP_AAPFR2022_Koubiti.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Keilhacker, M.; Watkins, M.L.; JET Team. D-T experiments in the JET tokamak. *J. Nucl. Mater.* **1999**, *266–269*, 1–13. [CrossRef]
2. Kim, H.; Sips, A.C.C.; Challis, C.D.; Keeling, D.; King, D.; Joffrin, E.; Szepesi, G.; Buchanan, J.; Horton, L.D.; Yuan, X.; et al. 1997 JET DT experiments revisited—comparative analysis of DD and DT stationary baseline discharges. *Nucl. Fusion* **2020**, *60*, 066003. [CrossRef]
3. Mailloux, J.; Abid, N.; Abraham, K.; Abreu, P.; Adabonyan, O.; Adrich, P.; Afanasev, V.; Afzal, M.; Ahlgren, T.; Aho-Mantila, L.; et al. Overview of JET results for optimizing ITER operation. *Nucl. Fusion* **2022**, *62*, 042026. [CrossRef]
4. Roth, J.; Tsitrone, E.; Loarer, T.; Philipps, V.; Brezinsek, S.; Loarte, A.; Counsell, G.F.; Doerner, R.P.; Schmid, K.; Ogorodnikova, O.V.; et al. Tritium inventory in ITER plasma-facing materials and tritium removal procedures. *Plasma Phys. Control. Fusion* **2008**, *50*, 103001. [CrossRef]
5. Drenik, A.; Alegre, D.; Brezinsek, S.; de Castro, A.; Kruezi, U.; Oberkofler, M.; Panjan, M.; Primc, G.; Reichbauer, T.; Resnik, M.; et al. Evaluation of the plasma hydrogen isotope content by residual gas analysis at JET and AUG. *Phys. Scr.* **2017**, *2017*, 014021. [CrossRef]
6. Klepper, C.C.; Hillis, D.L.; Bucalossi, J.; Douai, D.; Oddon, P.; Vartanian, S.; Colas, L.; Manenc, L.; Pégourié, B. Residual gas analysis for long-pulse, advanced tokamak operation. *Rev. Sci. Instrum.* **2010**, *81*, 10E104. [CrossRef]
7. Kubo, H.; Takenaga, H.; Sugie, T.; Higashijima, S.; Suzuki, S.; Sakasai, A.; Hosogane, N. The spectral profile of the Hα line emitted from the divertor region of JT-60U. *Plasma Phys. Control Fusion* **2014**, *56*, 025010.
8. Koubiti, M.; Marandet, Y.; Escarguel, A.; Capes, H.; Godbert-Mouret, L.; Stamm, R.; De Michelis, C.; Guirlet, R.; Mattioli, M. Analysis of asymmetric Dα spectra emitted in front of a neutralizer plate of the Tore-Supra ergodic divertor. *Plasma Phys. Control Fusion* **2002**, *44*, 261. [CrossRef]
9. Hey, J.; Chu, C.C.; Mertens, P.; Brezinsek, S.; Unterberg, B. Atomic collision processes with ions at the edge of magnetically confined fusion plasmas. *J. Phys. B* **2004**, *37*, 2543. [CrossRef]
10. Koubiti, M.; Sheeba, R.S. Spectral modeling of hydrogen radiation emission in magnetic fusion devices. *Atoms* **2019**, *7*, 23. [CrossRef]
11. Neverov, V.S.; Kukushkin, A.B.; Stamp, M.F.; Alkseev, A.G.; Brezinsek, S.; von Hellermann, M.; JET Contributors. Determination of diverter stray light in high-resolution main chamber Hα spectroscopy on JET-ILW. *Nucl. Fusion* **2017**, *57*, 016031. [CrossRef]
12. Neverov, V.S.; Kukushkin, A.B.; Kruezi, U.; Stamp, M.F.; Weisen, H.; JET Contributors. Determination of isotope ratio in the divertor of JET-ILW by high-resolution Hα spectroscopy: H-D experiment and implications for D-T experiment. *Nucl. Fusion* **2019**, *59*, 04601. [CrossRef]
13. Kajita, S.; Nishijima, D.; Fujii, K.; Akkermans, G.; van der Maiden, H. Application of multiple regression for sensitivity analysis of helium line emissions to the electron density and temperature in Magnum-PSI. *Plasma Phys. Control Fusion* **2021**, *63*, 055018. [CrossRef]
14. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Sickit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825.
15. Kajita, S.; Akkermans, G.; Fujii, K.; van der Maiden, H.; van de Sanden, M.C.M. Emission spectroscopy of He lines in high-density plasmas in Magnum-PSI. *AIP Adv.* **2020**, *10*, 025225. [CrossRef]
16. Nishijima, D.; Kajita, S.; Tynan, G.R. Machine learning predictions of electron density and temperature from He I line ratios. *Rev. Sci. Instrum.* **2021**, *92*, 023505. [CrossRef]
17. Gurung-López, S.; Gronke, M.; Saito, S.; Bonoli, S.; Orsi, A.A. zELDA: Fitting Lyman alpha line profiles using deep learning. *Mon. Not. R. Astron. Soc.* **2022**, *510*, 4525. [CrossRef]
18. Cianciosa, M.; Law, K.J.H.; Martin, E.H.; Green, D.L. Machine learning for analysis of atomic spectral data. *J. Quant. Spectrosc. Radiat. Transf.* **2020**, *240*, 106671. [CrossRef]
19. MacBride, C.D.; Jess, D.B.; Grant, S.D.T.; Khomenko, E.; Keys, P.H.; Stangalini, M. Accurately constraining velocity information from spectral imaging observations using machine learning techniques. *Phil. Trans. R. Soc. A.* **2021**, *379*, 20200171. [CrossRef]
20. Ren, T.; Modest, M.F.; Fateev, A.; Sutton, G.; Zhao, W.; Rusu, F. Machine learning applied to retrieval of temperature and concentration distributions from infrared emission measurements. *Appl. Energy* **2019**, *252*, 113448. [CrossRef]
21. Samuell, C.M.; Mclean, A.G.; Johnson, C.A.; Glass, F.; Jaervinen, A.E. Measuring the electron temperature and identifying plasma detachment using machine learning and spectroscopy. *Rev. Sci. Instrum.* **2021**, *92*, 043520. [CrossRef] [PubMed]
22. Skinner, C.H.; Ramsey, A.T.; Johnson, D.W.; Diesso, M. *Tritium Diagnostics by Balmer-Alpha Emission*; Princeton University: Princeton, NJ, USA, 1993; PPPL Report PPPL-2878. Available online: https://www.osti.gov/servlets/purl/6748245 (accessed on 15 August 2022). [CrossRef]
23. Skinner, C.H.; Stotler, D.P.; Adler, H.; Ramsey, A.T. Spectroscopic diagnostics of tritium recycling in TFTR. *Rev. Sci. Instrum.* **1995**, *66*, 646. [CrossRef]
24. TensorFlow. Available online: https://www.tensorflow.org/guide (accessed on 15 August 2022).
25. Keras Optimizers (ADAM optimizer). Available online: https://keras.io/api/optimizers/adam/ (accessed on 15 August 2022).
26. Keras Python API for Machine-Learning Platform TensorFlow. Available online: https://keras.io/ (accessed on 15 August 2022).
27. Sickit-Learn. Available online: https://scikit-learn.org/stable/ (accessed on 15 August 2022).