Научном већу Института за физику Београд

Предмет: Молба за покретање поступка за избор у звање истраживач сарадник

Молим научно веће Института за физику у Београду да покрене мој избор у звање истраживач сарадник.

У прилогу достављам:

1. мишљење руководиоца лабораторије са предлогом чланова комисије за избор у звање;
2. стручну биографију;
3. преглед научне активности;
4. списак објављених радова;
5. потврду о одржаном позивном предавању;
6. потврду о уписаним докторским студијама;
7. копију диплома основних и мастер студија;
8. уверење о прихваћеној теми докторске дисертације;

Београд,
08.08.2023. године

С поштовањем,
Иван Трапарић
истраживач приправник

Мишљење руководиоца лабораторије о избору Ивана Трапарића у звање истраживач сарадник

Иван Трапарић је изабран у звање истраживач приправник фебруара 2021. године и запослен је на Институту за физику у Београду од априла 2021. године као истраживач приправник у Лабораторији за спектроскопију плазме и ласере. Ради на примени вештачке интелигенције у спектроскопији плазме под руководством др Миливоја Ивковића и др Маријане Гавриловић Божовић. С обзиром да испуњава све предвиђене услове у складу са Правилником о поступку, начину вредновања и квантитативном исказивању научно истраживачких резултата Министарства науке, технолошког развоја и иновација, сагласан сам са покретањем поступка за избор Ивана Трапарића у звање истраживач сарадник.

За састав комисије за избор Ивана Трапарића у звање истраживач сарадник предлажем:

1. др Миливоје Ивковић, научни саветник, Институт за физику у Београду
2. др Биљана Станков, научни сарадник, Институт за физику у Београду
3. др Маријана Гавриловић Божовић, доцент, Факултет инжењерских наука Универзитета у Крагујевцу

др Миливоје Ивковић
научни саветник Института за физику
руководилац Лабораторије за спектроскопију плазме и ласере

Београд,
08.08.2023. године

Биографија

Иван Трапарић рођен је у Требињу, БиХ, 14.9.1996. године где је завршио основну и средњу школу. На основне студије Физичког факултета Универзитета у Београду уписује се 2015. године на смер *Примењена и комп̄јӯтерска физика*. Основне студије завршава у редовном року 2019. године са средњом просечном оценом 9.43. Исте године уписује мастер студије на Физичком факултету Универзитета у Београду на смеру *Теоријска и експ̄еримент̄ална физика*. Мастер студије је завршио са средњом просечном оценом 10, а мастер рад одбранио са оценом 10. Мастер рад под насловом „*Вакуум ултраљубичаст̄а спект̄роскоп̄ија Лајманове серије јонизованоī ат̄ома хелијума*" је урађен у Лабораторији за спектроскопију плазме и ласере на Институту за физику у Београду, под руководством др Миливоја Ивковића.

На Институту за физику у Београду је запослен од априла 2021. као студент докторских студија у звању истраживач приправник. Током докторских студија, бавио се или се бави вакуум ултраљубичастом спектроскопијом електричних гасних пражњења, применом машинског учења и вештачке интелигенције у оптичкој емисионој спектроскопији плазме и унапређивањем метода и побољшања граница детекције појединих елемената у спектроскопији ласерски индукованих плазми. У последње време, бави се и применом вештачке интелигенције у вакуум ултраљубичастој спектроскопији фузионих плазми. У току основних студија је учествовао на две летње школе из области физике фузионих плазми. Прва летња школа организована је у Београду, у огранизацији *Фузионе образовне мреже* (ФОМ). Током ове летње школе, имао је прилику да путем интернета присуствује експериментима на токамаку GOLEM у Прагу, а тема истраживања била је експериментална физика *runaway* електрона. Друга летња школа је била на Институту за физику плазме Чешке академије наука и уметности у Прагу, где су рађени експерименти на токамаку COMPASS. Овде је имао прилику да две недеље учествује у истраживању експерименталне групе која се бави нестабилностима на ивици плазме.

До сада је објавио три рада у часописима са СЦИ листе, од којих је један објављен у врхунском међународном часопису (категорија М21, ИФ 5.102) а друга два у међународном часопису категорије М23 (ИФ 1.611 и 0.420). Свој досадашњи рад презентовао је на три међународне конференције (SPIG 2020, SPIG 2022 и SLSP 6 (Spectral Line Shapes in Plasmas)). На међународној конференцији *14th Serbian Conference on Spectral Line Shapes in Astrophysics* одржаној у Бајној Башти у јуну ове године је одржао позивно предавање у секцији *Spectral Line Research: New Frontiers*.

Преглед научне активности Ивана Трапарића

Научно – истраживачки рад Ивана Трапарића састоји се од експерименталног истраживања у области спектрокоспије ласерски индукованих плазми и начина побољшања осетљивости методе, као и примене вештачке интелигенције у спектроскопији плазме.

(1) Истраживање могућности примене вештачке интелигенције у спектроскопији плазме

У оквиру докторске тезе Иван Трапарић је испитивао могућност примене вештачке интелигенције у спектрокопији плазме. Анализирана је примена вештачке интелигенције за одређивање Штаркове полуширине спектралне линије, као и за квантитативну анализу и одређивање концентрације елемената у узорку помоћу спектроскопије ласерски индуковане плазме. На крају, вештачка интелигенција је искориштена за генерисање спектра у области меког X зрачења за услове који се могу наћи у центру плазме у стелератору ЛХД у Јапану.

(2) Истраживање повећања осетљивости мерења методом спектроскопије ласерски индуковане плазме

У оквиру овог дела свог научног рада, испитана је могућност повећања интензитета емитоване спектралне линије из ласерски произведене плазме. Испитивани методи су додано електрично пражњење у две конфигурације. Једна конфигурација представља брзо импулсно пражњење чији је тригер ласерски произведена плазма, а други метод је убацивање аблираног материјала у тињаво пражњење. Мета на којој је предложена метода тестирана је волфрам допиран са ренијумом, који се очекује у првом зиду будућих фузионих реактора.

Списак објављених радова

Радови у врхунским међународним часописима (категорија М21)

Tapalaga, I., Traparić, I., Trklja Boca, N. *et al. Stark spectral line broadening modeling by machine learning algorithms.* Neural Comput & Applic 34, 6349–6358 (2022). https://doi.org/10.1007/s00521-021-06763-4

Радови у међународним часописима (категорија М23)

NM Sakan, I Traparic, VA Sreckovic, M Ivkovic *The usage of perceptron, feed and deep feed forward artificial neural networks on the spectroscopy data: astrophysical & fusion plasmas* Contrib. Astron. Obs. Skalnaté Pleso 52 (2022). https://doi.org/10.31577/caosp.2022.52.3.97

Traparić, I., Ivković, M. *Determination of austenitic steel alloys composition using laser-induced breakdown spectroscopy (LIBS) and machine learning algorithms.* Eur. Phys. J. D 77, 30 (2023). https://doi.org/10.1140/epjd/s10053-023-00608-6

Neural Computing and Applications
https://doi.org/10.1007/s00521-021-06763-4

ORIGINAL ARTICLE

Stark spectral line broadening modeling by machine learning algorithms

Irinel Tapalaga[1] · Ivan Traparić[1,2] · Nora Trklja Boca[1] · Jagoš Purić[1] · Ivan P. Dojčinović[1]

Received: 9 February 2021 / Accepted: 11 November 2021

Abstract
Various types of electric fields contained in the laboratory and astrophysical plasma cause a Stark broadening of spectral lines in plasma. Therefore, a large number of spectroscopic diagnostics of laboratory and astrophysical plasma are based on experimental and theoretical studies of Stark broadening of spectral lines in plasma. The topic of the present investigation is the Stark broadening caused by free electrons in plasma and its dependence on certain atomic parameters using a new method based on the machine learning (ML) approach. Analysis of empirical data on atomic parameters was done by ML algorithms with more success that it was previously done by classical methods of data analysis. The correlation parameter obtained by artificial intelligence (AI) is slightly better than the one obtained by classical methods, but the scope of application is much wider. AI conclusions are applicable to any physical system while conclusions made by classical analysis are applicable only to a small portion of these systems. ML algorithms successfully identified quantum nature by analyzing atomic parameters. The biggest issue of classical analysis, which is infinite spectral line broadening for high ionization stages, was resolved by AI with a saturation tendency.

Keywords Machine learning · Stark broadening · Atomic data · Plasma physics

1 Introduction

One of the greatest challenges of the modern science is the processing of enormous amounts of data. Two primary goals in this field are how to learn from data and how to make data predictions [37]. Data science and machine learning (ML) have made tremendous progress in the last few decades. Statistical physics have a great contribution to the development and understanding methods in ML [6]. ML algorithms have become very important in the analysis of data in physics and related sciences. ML methods have been shown to be useful in different physical sciences: astrophysics, particle physics, chemical physics, condensed-matter physics, quantum physics. In astrophysics and particle physics experiments such as CMS and ATLAS at the LHC in CERN, as well as projects such as the Sloan Digital Sky Survey (SDSS), gives enormous amount of data measuring the particle collisions and properties of a billion stars and galaxies [37]. Object classification in astrophysics is very important task [3, 27, 46] whose successful solution enables easier selection of objects according to certain criteria and their further study. Use of ML algorithms enables the solution of numerous problems in the processing of astronomical data and enables the consideration of dependences and correlations that have not been observed before [25, 26, 44, 45]. In one of the most significant modern experiment: observation of the gravitational waves arises from the merger of a binary black hole, ML algorithms are used to clear the noise in the signals of gravitational waves [1, 57]. Large development in the domain of data science, as well as the strengthening of connection to quantum physics, enabled the development of our knowledge of the matter. Namely, atomic physics is a base for the creation of new molecules and advanced materials, as well. The theoretical basis of atomic physics is quantum mechanics. Quantum mechanics has provided a base for successful research of molecular

✉ Nora Trklja Boca
nora@ff.bg.ac.rs

[1] Faculty of Physics, University of Belgrade, Studentski trg 12, Belgrade 11000, Serbia

[2] Institute of Physics, University of Belgrade, Pregrevica 118, Belgrade 11080, Serbia

Published online: 16 January 2022

physics and condensed matter physics, as well as nuclear physics and particle physics. Consequently, the development of ML algorithms recently has been used to address fundamental questions in the domain of quantum physics [4, 49]. One of the big problems in quantum physics is the inability to solve the Schrödinger equation for multiparticle systems, i.e., inability to obtain appropriate wave functions. In quantum chemistry and chemical physics, ML algorithms are used for analysis and prediction of physical and chemical properties, chemical structures, optimization of reaction parameters and process conditions (a type of reagents, catalysts, concentration, time, temperature), prediction of new reaction design, maximization of the production rate of chemical reactions [2, 9, 18, 29, 59, 60]. In condensed-matter physics and material science, ML has been used to improve the calculation of material properties, as well as a modeling of properties of new materials [10, 12, 19, 43, 58]. Learning techniques can be used for the research of advanced materials: materials for memory devices, solar cells, batteries, sensors, nanoparticle catalysts, supercapacitors, superhard material, etc.

From interconnections between quantum physics and ML, scientific community has huge expectation [4, 8, 23, 24, 32, 33, 35, 38, 49, 54]. There are many current questions about entanglement classification [23], quantum state tomography [54], solving the quantum many-body problem [8, 24]. There are many interesting active ML models with the aim of creation of new quantum experiments [32, 33, 38], as well as quantum machine models [17]. Also, ML can be used to discover physical concepts from new experimental data [28]. ML approaches have been applied to atomic physics. The neural networks can be used as a representation of quantum states [6]. There is a lack of atomic parameters for heavy elements and high ionized atoms. ML can be applied to the classification of heavy atoms energy levels according to their electronic configurations [7, 41]. Learning techniques have been applied for the investigation of atomic processes, like ionization and radiation [6]. Analysis of spectra from stars and quasars, as well as laser-produced and fusion plasma, can be done with the help of ML [40, 45]. The knowledge of atomic parameters is a base for atomic processes modeling. The aim of this research work is to use machine learning algorithms for modeling Stark spectral line broadening.

Stark broadening of spectral lines is a tool for spectroscopic diagnostics of laboratory plasma, as well as astrophysical and fusion plasma. In astrophysics, stark line widths are used for analysis of stellar spectra, investigation of chemical abundances of elements in different stellar objects, opacity calculations. Recently, it has been shown that Stark broadening has a big influence on the uncertainties in the calculation of the solar opacity [34]. Spectral analysis has a very important role in the physics of fusion plasma, too. Part of the current research in this field is concentrated on the possibility of using various durable materials, (as Mo, Ti, Zr...) for tungsten alloying. During the operation of fusion machines, it is expected that a small amount of these materials would be found in the peripheral regions of confined plasma, because of the spattering process. Stark widths of these atoms and their ions are needed for a detailed spectral analysis and diagnostics.

Stark broadening of spectral lines of neutral atoms and ions is used in science for a number of problems in various physical conditions. Theoretical calculations of the Stark width values usually use one of the models given by Griem et al. [20]; Sahal-Bréchot et al. [47]; Griem [21]; Dimitrijević and Konjević [13]. Recent research indicates the importance and usefulness of searching for possible types of regularities in the framework of a Stark broadening investigation [16, 53, 56]. Still, existing tables with calculated and measured Stark widths have a big lack of data. There is a need for Stark widths data in the wide range of chemical elements, plasma temperature and electron densities. In this paper a correlation between Stark broadening and environment parameters, such as the ionization potential of the upper level of the corresponding transition, electron density and temperature, will be investigated using modern ML algorithms. If this method proves to be accurate enough, the process of calculating the value of stark widths will be significantly accelerated and facilitated.

2 Theoretical background of Stark line broadening

One of the primary deexcitation ways of excited atoms is photon emission. A cumulative signal obtained from a radiating medium (for example plasma) is a spectral line with a small frequency range. Each spectral line emitted by atoms in plasma has a finite frequency range represented as line width. Generally, there are four types of spectral line broadening in plasma. One is a natural line broadening, related with the uncertainty in the energy of the states involved in the transition profile, with line width negligible compared to the other broadening mechanisms. The second is instrumental line broadening which includes the influence of spectral device on a line profile. Doppler broadening is caused by a distribution of velocities of atoms or molecules and depends on plasma temperature. The pressure broadening, which generally depends on pressure (i.e., density of active species) and temperature, results from the interactions of the emitters with neighboring neutral particles (resonant and Van der Waals broadening) and ionized particles (Stark broadening). Natural and instrumental broadening are always present. The existence of other types

of spectral line broadening depends on the plasma conditions. Stark broadening is caused by the free charges which surround the emitters in plasma and produce the local electric field which affects the emission process, giving rise to shifts of the emission wavelengths or changes in the phase of the radiation. This is observed as a phenomenon of broadening and shift of the spectral lines [22]. This effect is determined by the intensity of the local electric field and it depends on the density of charged particles in the plasma (electrons and ions). The influence of free electrons on the line broadening in plasma is much more pronounced than the influence of ions. The topic of the present investigation is the Stark broadening caused by free electrons in plasma using new method based on ML approach.

The general formula for Stark width calculation in the impact approximation, which is appropriate to use in the overwhelming number of cases, is [22]:

$$\omega = N_e \left\langle v \cdot \left(\sigma_{if} + \sum_{i'} \sigma_{ii'} + \sum_{f'} \sigma_{ff'} \right) \right\rangle_{av} \tag{1}$$

The line has a Lorentzian profile whose width is ω, expressed in angular frequencies unit [rad/s]. N_e and v are electron density and velocity, respectively. The average is being over the velocity distribution of perturbing electrons. The cross sections $\sigma_{ii'}$ ($\sigma_{ff'}$) are for inelastic scattering on the initial (final) state of the line,

while σ_{if} is an effective elastic cross section to be calculated essentially from the difference of elastic scattering amplitudes f_i and f_f.

The formula proposed by Griem [22] is very complicated, it cannot be resolved exactly, so it is useful to use different approaches in the calculation. The regularity approach which correlates Stark width of spectral line, ω, expressed in [rad/s], electron density N_e, electron temperature T_e and positive value of electron binding energy on the upper level of the transition, expressed in [eV], is given by Purić and Šćepanović [42] (Eq. 2):

$$\omega = Z_e^k \cdot a \cdot N_e \cdot f(T_e) \cdot \chi^{-b} \tag{2}$$

where $Z_e = 1, 2, 3...$ for neutrals, singly charged ions, ... respectively and it represents the rest core charge of the ionized emitter and a, b and k are coefficients independent of electron concentration and ionization potential for a particular transition and the rest core charge of the emitter. In Eq. 2, stark width is expressed in radian per second and this is the only suitable unit to analyze the regularity of Stark broadening. If regularity analysis use wavelengths (λ) and line widths ($\Delta\lambda$) expressed in meters, Eq. 2 should be written as follows:

$$\Delta\lambda = \frac{Z_e^k \cdot a \cdot N_e \cdot f(T_e) \cdot \chi^{-b} \cdot \lambda^2}{2 \cdot \pi \cdot c} \tag{3}$$

In Eq. 3, every transition have its own wavelength, while the other parameters remain unchanged, so regularity can't be seen.

Atoms and ions with the same number of electrons form an isoelectronic sequence. It is expected that spectral series within an isoelectronic sequence show regularity behavior because a wide range of atomic/ionic parameters depend on the electron number. The results obtained by regularity studies have proven to be very precise and this approach has been used in previous papers of our group. In the last decade we have investigated Stark broadening regularities using Eq. 2 within spectral series of individual elements [14, 15, 30, 51, 52], within spectral series of individual isoelectronic sequences [53, 55, 56] and we published one paper with analysis of Stark line broadening regularities within two spectral series of isoelectronic sequences simultaneously: potassium and copper [16]. The present investigation goes one step further and analyses all elements for which there are available data needed for Stark broadening investigation, simultaneously, using machine learning approach. The aim is to find the best possible model which correlates Stark width of spectral line with all available parameters for transition of interest (atomic parameters and environmental parameters).

3 Dataset creation and data cleaning

In order to create our dataset we used two public repositories connected with atomic spectroscopy. First one is Stark B database [48], where the parameters of Stark broadening for different emitters are given. The features taken from this database are: chemical element, ionization stage, upper and lower level of spectral transitions, Stark broadening, the environment temperature and electron density in environment. In the available database, stark widths are expressed in angstroms (1 angstrom = 10^{-10} m). For analysis purpose, angstroms are converted in radian per second [53] (Eq. 4). In the physical sense, radian per second is unit related with energy.

$$\omega = \frac{2\pi \cdot c \cdot \Delta\lambda}{\lambda^2} \tag{4}$$

We also performed data analysis with Stark widths expressed in meters, using machine learning and results have not shown any meaningful trends. This is confirmation that radians per second are suitable unit for machine learning approach, too. When the difference between energy levels of the same multiplet is small compared to the distance to the next level linked by an allowed

transition, all the fine structure lines of the same multiplet have the same width and shift in Stark B database . In that case, the data are given of the multiplet only (w_{mult}) and for an average wavelength of the whole multiplet (λ_{mult}). The width value for a particular line (w_{line}) within a multiplet is obtained from:

$$w_{\text{line}} = \frac{w_{\text{mult}} \cdot \lambda_{\text{line}}^2}{\lambda_{\text{mult}}^2} \tag{5}$$

To ensure better results, we enriched features taken from Stark B database with ones taken from NIST Atomic Spectra database [31]: binding energy of both upper and lower transition levels, ground level energy, total angular momentum quantum number J of both upper and lower transition level, as well as principal n and orbital ℓ quantum numbers and total angular momentum J quantum numbers of upper and lower transition levels.

The algorithm of connecting those two databases to form our own works as described below. For every transition connected with certain chemical element, we take the electronic configuration of both upper and lower levels from Stark B database. Then, we look for that particular element in NIST database and compare the electronic configurations. If they match, then we take the binding energy of those levels, their principal quantum number n, orbital quantum number ℓ and total angular momentum quantum number J and finally the ionization energy of that atom. The ionization energy is needed for calculation of the so-called upper level ionization potential χ. For practical purposes, we replaced the chemical element name with its atomic number Z and its charge, obtained from periodic system of elements (i.e., instead of Ar+2, we used $Z = 10$ and charge $= +2$). For better understanding of the algorithm mentioned above, pseudo code is given in table Algorithm 1 while the complete code can be found at https://github.com/ivantraparic/StarkBroadeningMLApproach. After we completed the creation of database, it consisted of 54,236 successfully matched transitions for 53 different emitters.

Then we proceeded with data cleaning. Outliers were detected as those transitions where the energy of upper level is smaller than the energy of lower level, which is physically impossible, thus those lines were removed from the database. Next, we excluded transitions given for temperatures above 150,000 K and electron densities above $10^{18}\,\text{cm}^{-3}$, because we are currently not interested in making predictions for those plasma conditions. As a result, this dataset contains 53 emitters and 34,973 spectral lines and follows a normal distribution.

Algorithm 1 Creation of database used in this paper

```
for element in elements do
    charge, Z ← DetermineZAndCharge(element)
    elementNIST ← FindElementInNIST(element)
    ElectronTemperature ← electronic temperature from Stark B database
    ElectronDensity ← electron density from Stark B database
    StarkWidth ← stark broadening from Stark B database
    UpperLevels ← upper transition levels from Stark B for element
    LowerLevels ← lower transition levels from Stark B for element
    Levels ← levels configuration in NIST database for elementNIST
    Upperbindingenergy, Lowerbindingenergy, jupper, jlower, ni, li, nf, lf, χ ← Arrays
for saving found quantities from NIST database
    GroundLevel ← FindGroundLevel(elementNIST)
    for lowerlevel in LowerLevels do
        for level in Levels do
            if lowerlevel == level then
                Lowerbindingenergy, nf, lf, jlower ← binding energy
 and associated quantum numbers for that lower level taken from NIST
                upper level ionization potential χ ← GroundLevel - Lowerbingingenergy
                break
    for upperlevel in UpperLevels do
        for level in Levels do
            if upperlevel == level then
                Upperbindingenergy, ni , li, jupper ← binding energy
 and associated quantum numbers for that upper level taken form NIST
                break
    if length(Lowerbindingenergy) != length(Upperbindingenergy) then
        Error
        return
    else
        for i != length(Lowerbindingenergy) do
            Z, ElectronTemperature, ElectronDensity, charge, χ, GroundLevel,
 Upperbindingenergy, jupper, Lowerbindingenergy, jlower, ni, li, nf, lf, StarkWidth
← Insert line i in database
```

Table 1 List of ML algorithms and parameters

Model	Parameters
Linear regression	Normalize: [True, False]
Decision tree regressor	max_depth [3, 5, 10]
Random forest regressor	n_estimators [5, 10, 15, 100]
Gradient boosting regressor	max_depth [3, 5, 10], n_estimators [100, 150, 200]

Table 2 List of ML algorithms and their final score

Model	Best R^2	Parameters
Linear regression	0.38	Normalize: False
Decision tree regressor	0.92	max_depth = 10
Gradient boosting regressor	0.94	max_depth = 10, n_estimators = 150

It consisted of 15 columns, 14 of those were our features (atomic number Z, electron temperature, electron density, charge, lower level ionization potential χ, ground level energy, lower level energy, lower level J, upper level energy, upper level J, principal quantum number of lower level n_f, orbital quantum number of lower level ℓ_f, principal quantum number of upper level n_i, orbital quantum number of upper level ℓ_i), and 15$^{\text{th}}$ column was our target value ω.

4 Model creation and training

For model creation and training, we used public Python package Sci-kit learn. We created four models, every being Pipeline with two steps. In each object of Pipeline class, the first step was data scaling using StandardScaler, and in second step we made our predictions with defined model. Considered models were: Linear Regression, Decision Tree Regressor, Random Forest Regressor and Gradient Boosting Regressor. We split the dataset into training and test dataset using train_test_split method, leaving 25% of the data for testing. To find the best model out of four considered, and best parameters for that model, we used ShuffleSplit combined with GridSearchCV. In ShuffleSplit object we set number of splits to 5, and we left 30% of data for testing. In GridSearchCV object we set cross validation to ShuffleSplit object. The values for parameters of models used in GridSearchCV are presented in table 1. Other parameters of the model remained at their default values.

To rank the performance of models, we used best Coefficient of Determination, R^2, value obtained after

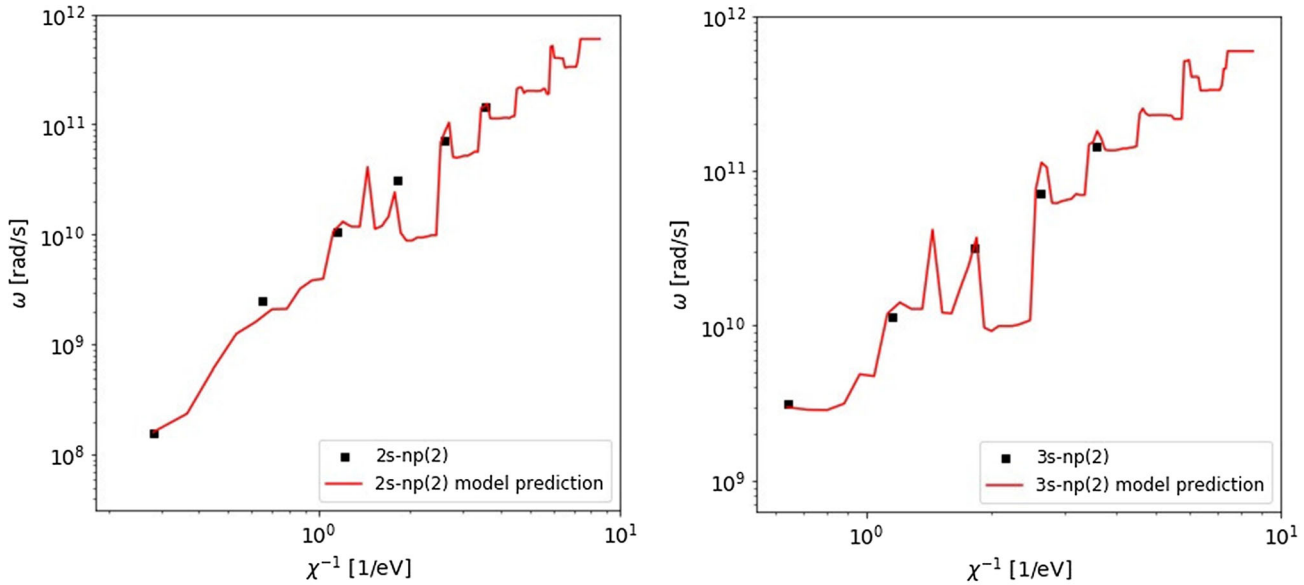


Fig. 1 Stark widths regularities within 2s-np i 3s-np spectral series of Li I (T = 30,000 K, $N_e = 10^{20}\,\mathrm{m}^{-3}$)

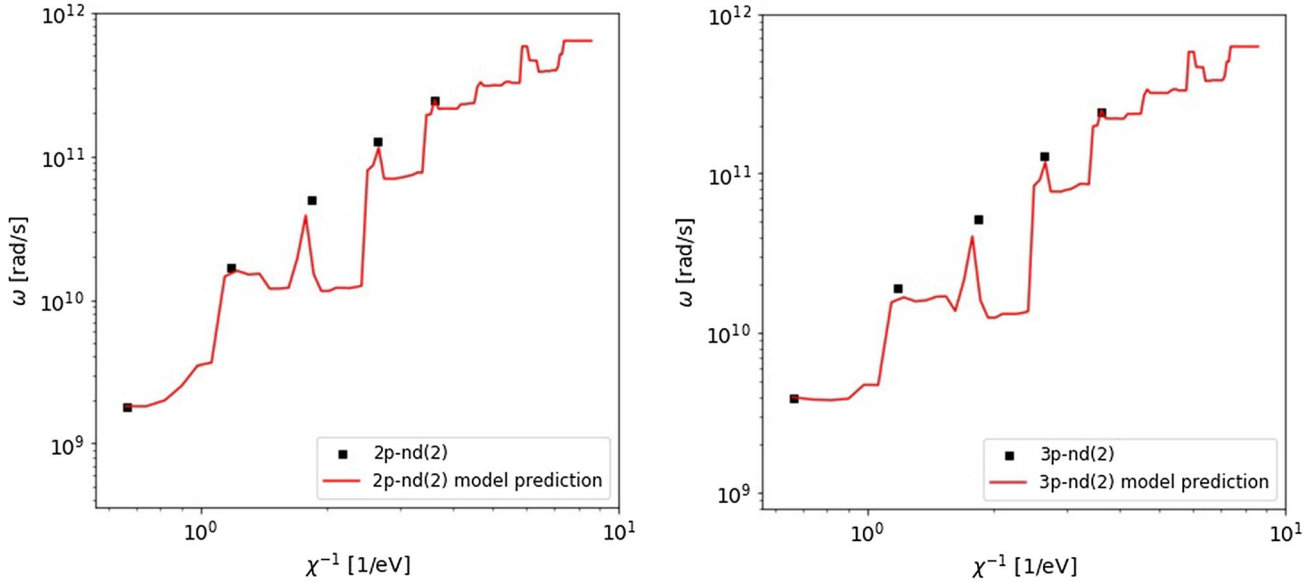


Fig. 2 Stark widths regularities within 2p-nd and 3p-nd spectral series of Li I (T = 30,000 K, $N_e = 10^{20}$ m^{-3})

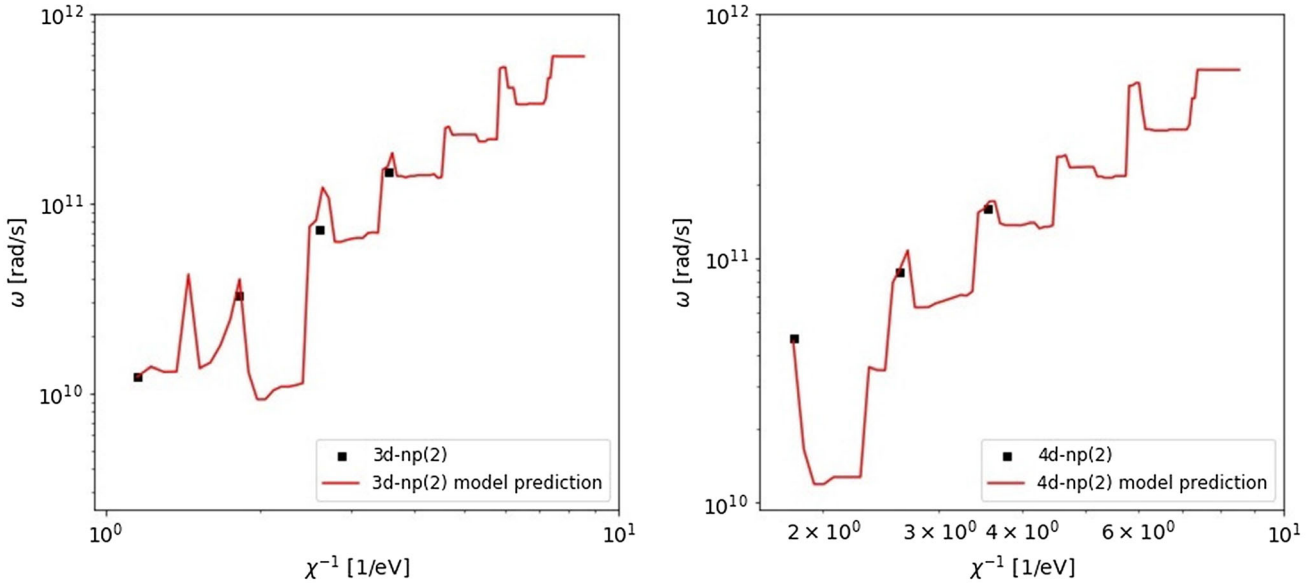


Fig. 3 Stark widths regularities within 3d-np and 4d-np spectral series of Li I (T = 30,000 K, $N_e = 10^{20}$ m^{-3})

GridSearchCV algorithm finished. We have taken the parameters that the algorithm used to score that particular R^2. As a result, we got that the best R^2 value was for Random Forest Regressor having R^2 = 0.95 for n_estimators = 100. The results of other model is given in table 2.

So, our winning model after performing hyper parameter tuning using GridSearchCV was Random Forest Regressor with number of estimators set to 100. Random Forest is a learning method that operates by constructing a large number of decision trees during the training process [5]. It is simple to use and shows high performance for a wide variety of tasks, making it one of the most popular ML algorithms in different sciences. Random forests are an effective tool in predicting new data, in our case new atomic parameters. It should be emphasized that Breimans paper [5] is cited more about 30,000 times (Web of Science: 36.234, CrossRef: 27.683). In order to check overfitting of the winning model, we did R^2 score check of the model on both training and test datasets. Training R^2 score was 0.98, and test score was $R^2 = 0.95$, so we were sure our model is not overfitting the data.

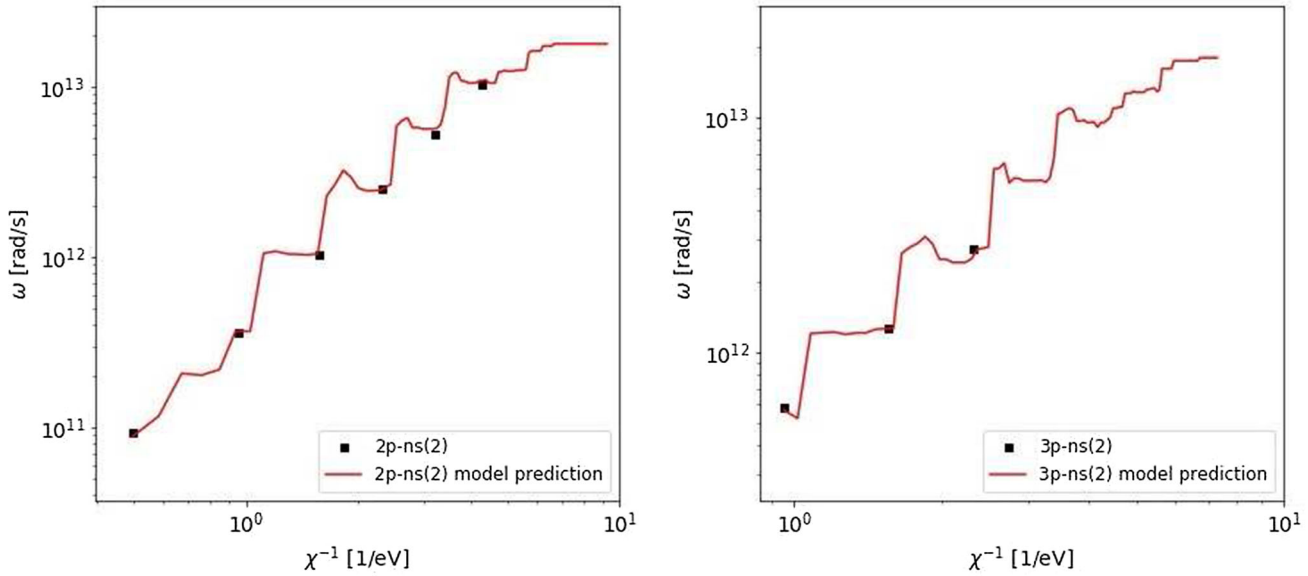


Fig. 4 Stark widths regularities within 2p-ns and 3p-ns spectral series of Li I (T = 30,000 K, $N_e = 10^{22}\,\mathrm{m}^{-3}$)

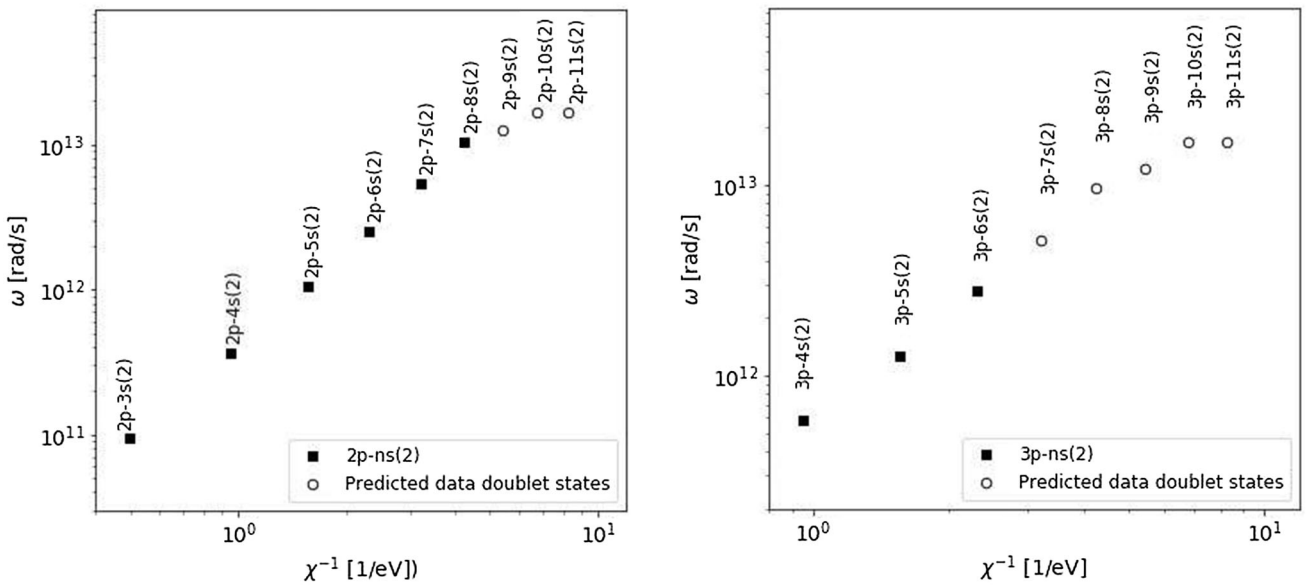


Fig. 5 Stark widths predictions for 2p-ns and 3p-ns spectral series of Li I (T = 30,000 K, $N_e = 10^{22}\,m^{-3}$)

5 Results

The Random Forest model is used to calculate Stark broadening data for spectral series within neutral lithium Li I. Calculated stark widths (red lines) for transitions within analyzed series are represented with existing known values of Stark widths data at the same graphs.

Figure 1 shows the dependence of the Stark width (ω) on the reciprocal value of the electron binding energy at the upper level of the transition ($\chi^{-}1$) for 2s-np and 3s-np transitions within lithium atom at a temperature of T = 30,000 K and electron concentration $N_e = 10^{20}\,\mathrm{m}^{-3}$.

Figures 2 and 3 show the change in Stark width under the same conditions (temperature $T = 30,000\,\mathrm{K}$ and electron concentration $N_e = 10^{20}\,\mathrm{m}^{-3}$), but for 2p-nd and 3p-nd transitions, as well as 3d-np and 4d-np. A very good description of the atomic structure of lithium, i.e., the values of atomic parameters, can be observed. Interestingly, points 2s-2p and 2p-3d were omitted in our previous analysis [15], while the ML algorithm includes them in the overall analysis and gives excellent agreement with the Stark width value.

Of special importance is the possibility of obtaining the value of stark widths at higher energy levels, for which

these data are not quantitatively calculated. Figure 4 shows the change of Stark widths for the 2p-ns and 3p-ns transitions at a temperature of T = 30,000 K and electronic concentration $N_e = 10^{22}\,\mathrm{m}^{-3}$. In Fig. 5, the conditions are the same, but the initial values of the Stark width and the values obtained by the ML algorithm are given for higher energy levels of lithium atom. A slight saturation of the values of Stark parameters at higher energy levels can be observed.

The functional dependence obtained using the ML algorithm describes the quantum structure of the energy levels of lithium atoms. From the model lines (red lines), it can be concluded that the model successfully (within the error) indicates the quantum nature of atomic transitions and that other results do not make physical sense, but only jumps.

6 Conclusion

Analysis of spectral data on Stark broadening for 53 different emitters and 34973 lines by ML algorithms was done with more success than it was previously done by classical methods of data analysis. Random forest has scored an average of $R^2 = 0.95$ which makes it an excellent choice for Stark broadening calculations. The correlation parameter obtained by AI is slightly better than the one obtained by classical methods of Stark broadening analysis, but the scope of application is much wider. AI conclusions are applicable to any physical system while conclusions made by classical analysis are applicable only to a small portion of these systems, mostly to ions with low ionization stage. This improves the quality of predictions and enhance a broader usability of results. In fact, these results can be used for any transition and any environment without any restrictions. ML algorithms successfully identified quantum nature by analyzing Stark broadening parameters which can not be done with similar analyses that used classical methods and obtained linear correlation. The biggest issue of the classical analysis is infinite spectral line broadening for high ionization stages and it was successfully resolved by AI with a saturation tendency.

The process of calculating the values of Stark widths, which is used in science for a number of problems in various physical conditions in spectroscopic diagnostics of laboratory plasma, as well as astrophysical and fusion plasma, is significantly accelerated and facilitated with new method based on ML and proposed in the present paper. Using our new proposed model, Stark databases can be significantly improved. For example, Stark broadening calculations can be made for some spectral transitions within W, Ti, Mo and Zr atoms, which are common in nuclear fusion diagnostics and of interest for spectral analysis in fusion physics [39], as titanium, molybdenum and zirconium are used as alloying materials for tungsten [11, 50]. Lines of Ti are used for astrophysics diagnostics, too [36]. Despite their significance, there is a very big lack of Stark data for these atoms and their ions. There is a lot of missing energy data for higher energy levels for W, Ti, Mo and Zr atoms and their ions. For example, there is no precise value of energy for 6p level for Ti II, so there is no calculated Stark data for transition for which this level is the closest perturbing level. Although very rich in data, NIST database does not have energy values data for all possible excited states of atoms. With standard known methods for Stark width calculation, it is not possible to calculate Stark widths for levels for which energy values of the closest perturbing levels are missing, but ML algorithms enable calculation in these situations, too. In next step, ML predictions and analytic calculations will be compared and ML technique will be used for atomic parameters analyses. Special attention will be paid to spectral lines that are important for fusion and astrophysical research.

Funding The Funding was provided by Ministarstvo Prosvete, Nauke i Tehnološkog Razvoja. This work is financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Abbott BP, Abbott R, Abbott TD, Abernathy MR, Acernese F, Ackley K, Adams C et al (2016) Observation of gravitational waves from a binary black hole merger. Phys Rev Lett 116:061102
2. Balabin RM, Lomakina EI (2009) Neural network approach to quantum-chemistry data: accurate prediction of density functional theory energies. J Chem Phys 131:074104
3. Ball NM, Brunner RJ (2010) Data mining and machine learning in astronomy. Int J Mod Phys 19(7):1049–1106
4. Bharti K, Haug T, Vedral V, Kwek LC (2020) Machine learning meets quantum foundations: a brief survey. AVS Quantum Sci 2:034101
5. Breiman L (2001) Random Forests. Mach Learn 45:5–32
6. Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, Zdeborova L (2019) Machine learning and the physical sciences. Rev Mod Phys 91:045002
7. Cao X, Liu H, Chen N (1997) Classification of Cm I energy levels using PCA-BPN and PCA-NLM. Chem Phys 220:289–297
8. Carleo G, Troyer M (2017) Solving the quantum many-body problem with artificial neural networks. Science 355:602–606

9. Cova TFGG, Pais AACC (2019) Deep learning for deep chemistry: optimizing the prediction of chemical patterns. Front Chem 7:809
10. Carrasquilla J, Melko RG (2017) Machine learning phases of matter. Nat Phys 13:431–434
11. Cui Z, Zhang X, Liu Q, Li H, Liu Y, Liu H, Wang X, Huang J, Liu H, Cheng J, Li M (2020) A first-principles study of the mechanical and thermodynamic properties of WTi, WV, W2Zr, WVTi, WVZr alloys. Fusion Eng Design 152:111451
12. Deringer VL, Caro MA, Csanyi G (2019) Machine learning interatomic potentials as emerging tools for materials science. Adv Mater 1902765:1–16
13. Dimitrijević MS, Konjević N (1980) Stark widths of doubly- and triply-ionized atom lines. JQSRT 24:451
14. Dojčinović IP, Tapalaga I, Purić J (2012) Stark parameter regularities of neutral helium lines within different spectral series. Mon Not R Astron Soc 419(1):904–912
15. Dojčinović IP, Tapalaga I, Purić J (2013) Stark-width regularities of neutral lithium lines within different spectral series. Mon Not R Astron Soc 429(3):2400–2406
16. Dojčinović IP, Trklja N, Tapalaga I, Purić J (2019) Investigation of Stark line broadening within spectral series of potassium and copper isoelectronic sequences. Mon Not R Astron Soc 489(3):2997–3002
17. Dunjko V, Wittek P (2020) A non-review of quantum machine learning: trends and explorations. Quantum Views 4:32
18. Gastegger M, Behler J, Marquetand P (2017) Machine learning molecular dynamics for the simulation of infrared spectra. Chem Sci 8:6924–6935
19. Gavezzotti A (2012) Computational studies of crystal structure and bonding. Top Curr Chem 315:1–32
20. Griem HR, Baranger M, Kolb AC, Oertel G (1962) Stark Broadening of Neutral Helium Lines in a Plasma. Phys Rev 125:177
21. Griem HR (1968) Semiempirical Formulas for the Electron-Impact Widths and Shifts of Isolated Ion Lines in Plasmas. Phys Rev 165:258
22. Griem HR (1974) Spectral line broadening by plasmas. Academic Press, New York
23. Harney C, Pirandola S, Ferraro A, Paternostro M (2020) Entanglement classification via neural network quantum states. New J Phys 22:045001
24. Hartmann MJ, Carleo G (2019) Neural-network approach to dissipative quantum many-body dynamics. Phys Rev Lett 122:250502
25. Hezaveh YD, Levasseur LP, Marshall PJ (2017) Fast automated analysis of strong gravitational lenses with convolutional neural networks. Nature 548:555–557
26. Huerta EA, Allen G, Andreoni I, Antelis JM, Bachelet E, Berriman GB, Bianco FB et al (2019) Enabling real- time multi-messenger astrophysics discoveries with deep learning. Nat Rev Phys 1:600–608
27. Ishida EEO, Beck R, González-Gaitán S, de Souza RS, Krone-Martins A, Barrett JW, Kennamer N et al (2019) Optimizing spectroscopic follow-up strategies for supernova photometric classification with active learning. Mon Not R Astron Soc 483:2–18
28. Iten R, Metger T, Wilming H, del Rio L, Renner R (2020) Discovering physical concepts with neural networks. Phys Rev Lett 124:010508
29. Janet JP, Chan L, Kulik HJ (2015) Accelerating Chemical discovery with machine learning: simulated evolution of spin crossover complexes with an artificial neural network. J Phys Chem Lett 9:1064–1071
30. Jevtić D, Dojčinović IP, Tapalaga I, Purić J (2012) Stark width regularities of neutral potassium lines within different spectral series. Bull Astr Soc India 40:151–160
31. Kramida A, Ralchenko Y, Reader J, NIST ASD Team (2019) NIST Atomic Spectra Database (version 5.7.1), [Online]. Available: https://physics.nist.gov/asd [Tue Mar 24 (2020) National Institute of Standards and Technology. Gaithersburg, MD
32. Krenn M, Gu X, Zeilinger A (2017) Quantum experiments and graphs: Multiparty states as coherent superpositions of perfect matchings. Phys Rev Lett 119:240403
33. Krenn M, Malik M, Fickler R, Lapkiewicz R, Zeilinger A (2016) Automated search for new quantum experiments. Phys Rev Lett 116:090405
34. Krief M, Feigel A, Gazit D (2016) Line broadening and the solar opacity problem. Astrophy J 824(2):98
35. Luchnikov IA, Vintskevich SV, Grigoriev DA, Filippov SN (2020) Machine learning non-Markovian quantum dynamics. Phys Rev Lett 124:140502
36. Manrique J, Aguilera JA, Aragon C (2016) Experimental Stark widths and shifts of Ti II spectral lines. Mon Not R Astron Soc 462(2):1501–1507
37. Mehta P, Bukov M, Wanga CH, Day AGR, Richardson C, Fisher CK, Schwab DJ, colleagues, (2019) A high-bias, low-variance introduction to Machine Learning for physicists. Phys Rep 810:1–124
38. Melnikov AA, Nautrup HP, Krenn M, Dunjko V, Tiersch M, Zeilinger A, Briegel HJ (2018) Active learning machine learns to create new quantum experiments. Proc Natl Acad Sci USA 115(6):1221–1226
39. Mišković ova J, Anguš M, van der Meiden H, Veis P (2020) Selection of molybdenum lines by quantitative analysis of molybdenum-zirconium-titanium alloy by CF-LIBS for future fusion applications. Fusion Eng Des 153:111488
40. Osterheld AL, Morgan WL, Larsen JT, Young BKF, Goldstein WH (1994) Analysis of spectra from laser produced plasmas using a neural network. Phys Rev Lett 73(11):1505–1508
41. Peterson KL (1991) Classification of Cm II and Pu I energy levels using counterpropagation neural networks. Phys Rev A 44:126–138
42. Purić J, Šćepanović M (1999) General regularities of stark parameters for ion lines. Astroph J 521:490
43. Purja Pun GP, Batra R, Ramprasad R, Mishin Y (2019) Physically informed artificial neural networks for atomistic modeling of materials. Nat Commun 10:2339
44. Reis I, Baron D, Shahaf S (2019) Probabilistic random forest: a machine learning algorithm for noisy data sets. Astronom J 157(16):1–12
45. Reis I, Poznanski D, Baron D, Zasowski G, Shahaf S (2018) Detecting outliers and learning complex structures with large spectroscopic surveys - a case study with APOGEE stars. Mon Not R Astron Soc 476:2117–2136
46. Rohde DJ, Drinkwater MJ, Gallagher MR, Downs T, Doyle MT (2005) Applying machine learning to catalogue matching in astrophysics. Mon Not R Astron Soc 360:69–75
47. Sahal-Bréchot S (1969) Impact theory of the broadening and shift of spectral lines due to electrons and ions in a plasma. Astron Astrophys 1:91
48. Sahal-Bréchot S, Dimitrijević MS, Moreau N (2020) STARK-B database, [online]. Available: http://stark-b.obspm.fr [February 20, 2020]. Observatory of Paris, LERMA and Astronomical Observatory of Belgrade
49. Sarma SD, Deng DL, Duan LM (2019) Machine learning meets quantum physics. Phys Today 72(3):48–54
50. Snead LL, Hoelzer DT, Rieth M, Nemith AA (2019) Refractory Alloys: Vanadium, Niobium, Molybdenum, Tungsten. Struct

Alloys for Nucl En App 585–640, Chapter 13. https://www.sciencedirect.com/science/article/pii/B9780123970466000137
51. Tapalaga I, Dojčinović IP, Purić J (2011) Stark width regularities within magnesium spectral series. Mon Not R Astron Soc 415:503
52. Tapalaga I, Dojčinović IP, Milosavljević MK, Purić J (2012) Stark Width Regularities within Neutral Calcium Spectral Series. PASA 29:20
53. Tapalaga I, Trklja N, Dojčinović IP, Purić J (2018) Stark width regularities within spectral series of the lithium isoelectronic sequence. Mon Not R Astron Soc 474:5479
54. Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R, Carleo G (2018) Neural-network quantum state tomography. Nat Phys 14:447
55. Trklja N, Tapalaga I, Dojčinović IP, Purić J (2018) Stark widths regularities within spectral series of sodium isoelectronic sequence. New Astr 59:54
56. Trklja N, Tapalaga I, Dojčinović IP, Purić J (2019) Stark Widths Regularities Within: ns-np, np-ns, np-nd, nd-np and nd-nf Spectral Series of Potassium Isoelectronic Sequence. Atoms 7(4):99
57. Wei W, Huerta EA (2020) Gravitational wave denoising of binary black hole mergers with deep learning. Phys Lett B 800:135081
58. Westerhout T, Astrakhantsev N, Tikhonov KS, Katsnelson MI, Bagrov AA (2020) Generalization properties of neural network approximations to frustrated magnet ground states. Nat Commun 11:1593
59. Wu J, Shen L, Yang W (2017) Internal force corrections with machine learning for quantum mechanics/molecular mechanics simulations. J Chem Phys 147:161732
60. Zhang P, Shen L, Yang W (2019) Solvation free energy calculations with quantum mechanics / molecular mechanics and machine learning models. Phys Chem B 123(4):901–908

Eur. Phys. J. D (2023)77:30
https://doi.org/10.1140/epjd/s10053-023-00608-6

THE EUROPEAN PHYSICAL JOURNAL D

Regular Article – Plasma Physics

Determination of austenitic steel alloys composition using laser-induced breakdown spectroscopy (LIBS) and machine learning algorithms

Ivan Traparić[a] and Milivoje Ivković[b]

Institute of Physics Belgrade, Pregrevica 118, Belgrade 11080, Serbia

Received 5 November 2022 / Accepted 1 February 2023

Abstract. In this paper, the determination of composition of certified samples of austenitic steel alloys was done by combining laser-induced breakdown spectroscopy (LIBS) technique with machine learning algorithms. Isolation forest algorithm was applied to the MinMax scaled LIBS spectra in the spectral range form (200–500) nm to detect and eject possible outliers. Training dataset was then fitted with random forest regressor (RFR) and Gini importance criterion was used to identify the features that contribute the most to the final prediction. Optimal model parameters were found by using grid search cross-validation algorithm. This was followed by final RFR training. Results of RFR model were compared to the results obtained from linear regression with $\mathcal{L}^2$ norm and deep neural network (DNN) by means of R^2 metrics and root-mean-square error. DNN showed the best predictive power, whereas random forest had good prediction results in the case of Cr, Mn and Ni, but in the case of Mo, it showed limited performance.

1 Introduction

The structural materials of fusion reactors are subjected to thermal, mechanical, chemical, and radiation loads. Due to their excellent manufacturability, good mechanical properties, welding ability, and corrosion resistance, austenitic stainless steels were chosen as structural reference material for ITER [1]. In addition, entire vacuum vessel of LHD stellarator in Japan is made of austenitic steel [2], and to diagnose the composition of the deposits on the fusion reactor's first wall, test targets made of austenitic steel (AISI 316 L) were settled at ten positions on the first wall [3]. Laser-induced breakdown spectroscopy (LIBS) is one of the emerging analytical technique that is non-destructive, easy to use and requires little to no preparation of the sample [4]. Therefore, it represents a great tool for the analysis of the composition of austenitic steel samples. There are two main approaches to the LIBS analysis, namely standard calibration method and calibration—free method [5]. In the method where calibration curve is constructed, a connection between one integrated line intensity and known concentrations is established, thus enabling the determination of unknown concentration. This method is by far the most used one. Alternatively, one can assume local thermodynamic equilibrium (LTE) in plasma and use Saha–Boltzmann equation to obtain plasma temperature and density, and from this the unknown concentrations regardless of the matrix effect. Machine learning algorithms have been successfully applied in analysis of Raman spectra, NIR and THz spectroscopy, vibrational spectroscopy, fusion plasma spectroscopy, etc., just to name a few [6–10]. In recent years, to speed up the analysis of LIBS spectra, machine learning methods are being used intensively [11–14]. These methods involve the usage of principal component analysis (PCA) for dimensionality reduction, support vector machine (SVM) for classification purposes and partial least squares regression (PLS) for multivariate regression problems [15]. Also, for classification or regression problems, many authors applied back propagation neural networks (BPNN) or convolutional neural networks (CNN) to the LIBS spectra in order to perform quantitative analysis of different samples [16–20]. Other regression algorithms, like random forest regression (RFR), have also been widely used [21–24]. Random Forest was constructed and reported by Breiman [25], and it is based on the ensemble of decision threes, where the decision or prediction is made

Physics of Ionized Gases and Spectroscopy of Isolated Complex Systems: Fundamentals and Applications. Guest editors: Bratislav Obradović, Jovan Cvetić, Dragana Ilić, Vladimir Srećković, Sylwia Ptasinska.

[a] e-mail: traparic@ipb.ac.rs (corresponding author)
[b] e-mail: ivke@ipb.ac.rs

Published online: 22 February 2023

by the majority prediction. This algorithm was previously applied on steel spectra by Zhang et al. [26] where they showed that this regression could be applied for the determination of composition of steel alloys. Later, Zhang with his collaborators used BPNN combined with SelectKBest algorithm for feature selection to trace minor elements in steel samples [27]. Liu and his coworkers also used random forest, combined with permutation importance feature selection to train and predict the composition of steel alloys [28]. Gini importance criteria was also used previously in combination with random forest on classification problems [29,30], but here we are applying it to regression problem.

In this paper, we will consider three algorithms, random forest, linear regression with $\mathcal{L}^2$ norm and deep neural network (DNN) to predict steel samples composition. Instead of making our own database, we will use the dataset published at the LIBS 2022 conference site [31] and record our own test dataset under similar conditions to check how much these small differences affect the final model performance. Idea to use RF algorithm is twofold. On the one end, it is able to catch nonlinear phenomena in the data, on the other end to see to what extent we can use already implemented Gini importance criteria within RF to make good regression model. Although simple neural networks have yielded good analytical prediction in the past, in general, they are hard to train (better said, it is not easy to find most favorable architecture), so we wanted to see how close RF predictions are going to be with respect to DNN.

The paper is organized as follows: In the first section, a brief introduction and overview of previous results is given. In Sect. 2, the experimental setup and sample preparation is described. Section 3 gives the detail description of applied methodology and data preprocessing, while the results are given in Sect. 4. Finally, we gave the conclusion of this work in Sect. 5.

2 Experimental setup and sample preparation

Experimental setup is shown in Fig. 1.

The setup is a classical LIBS setup consisting of Quantel Q switched neodymium-doped yttrium aluminum garnet (Nd:YAG) laser having pulse width of 6 ns, repetition rate of 10 Hz, pulse energy of 96 mJ and operated at fundamental wavelength $\lambda = 1064$ nm. Laser beam was reflected from 45° angle mirror M and focused via lens L onto a target mounted on a x–y micrometric moving stage by a lens of focal length $f = 11$ cm. Light emitted from plasma was collected using a fiber optic cable with collimator having a focal length of $f_{fc} = 4.4$ cm and directed onto the 50 μm width entrance slit of Mechelle 5000 spectrograph that can record spectra from 200 to 950 nm. As a detector, we used Andor iStar ICCD camera (model DH734, 1024 × 1024 pixels) cooled to −15 °C. Camera was triggered with a photodiode and gated by usage of Stanford Research digital delay unit (model DG535). Delay from laser pulse was set to 0.6 μs and the gate width was set to 50 μs.

Steel samples used in this work were AISI steels with certified composition from National Bureau of Standards (NBS, today NIST), whose elemental composition is given in Table 1.

Sample mentioned above, austenitic steel AISI 316 L lies in between these tested models (concentrations of main elements: Cr 17%, Ni 12%, Mo 2% and Mn 2%). Each sample was firstly polished by sandpaper 200, followed by polishing it with sandpaper 600. In front of laser beam, external shutter was placed, coupled with laser pulse counter. Counter was set to 16 counts, as it is a binary counter, and after 16 pulses, the shutter is closed for another 16 pulses. This represents one acquisition of the spectra. For each sample, we recorded 22 spectra from different places on the target, and each spectra is a result of averaging 20 acquisitions on the same place (this gives 320 individual laser shots per place on the target). To further improve and increase signal, electrical gain of the camera was set to 80 (on the scale of 0–255).

3 Methodology and data preprocessing

Database used in this paper was downloaded from LIBS 2022 website [31]. This database consists of a spectra of 42 different steel samples, and for each sample, a 50 single-shot spectra were taken. This gives in total a database of 2100 spectra samples divided into 40,002

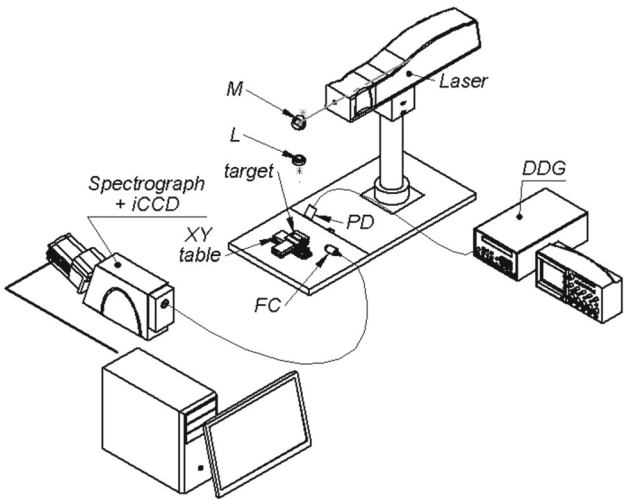


Fig. 1 Experimental setup. Laser (Quantel, $\lambda = 1064$ nm, pulse width 6 ns, peak energy 96 mJ) was focused via lens L onto the movable target and plasma spectrum was recorded by Andor iStar iCCD camera mounted on Echelle spectrograph. Camera gating was done by Stanford Research Digital Delay Generator (DDG, model 535) and triggered by photodiode (PD). Mirror M and lens L are integrated within a laser head, which was not drawn on this figure

Table 1 Steel alloy certified composition

Steel number	Steel type	Cr	Mn	Mo	Ni
443	Cr18.5–Ni9.5	18.5	3.38	0.12	9.4
445	AISI 410	13.31	0.77	0.92	0.28
446	AISI 321	18.35	0.53	0.43	9.11
447	AISI 309	23.72	0.23	0.053	13.26

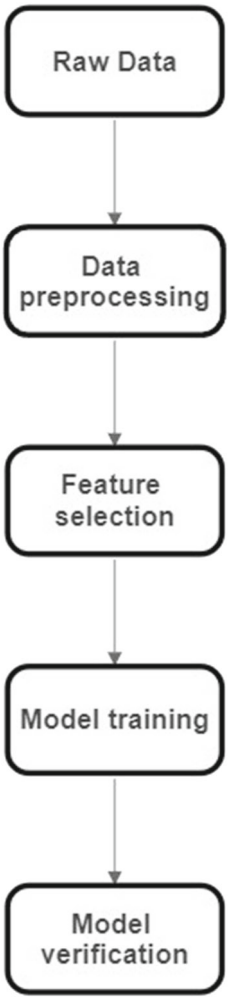


Fig. 2 Flowchart of procedures taken in this work

columns (each column corresponds to one wavelength). The flow diagram of our methodology is given in Fig. 2. For machine learning part of this work, we used python public repository scikit-learn.

3.1 Data preprocessing

Firstly, we restricted our dataset to the spectral range between 200 and 500 nm, as this is the spectral area where the most emission lines of metals of interest can be found. It is worth mentioning that all training dataset spectra were not intensity corrected. Therefore, no intensity correction was done on the test dataset. In the spectra normalization step, two normalizations were tried to later adopt the best one, and those were total spectral area normalization, and standard normal variate (SNV) normalization. First one is clear, whereas SNV normalization represents a spectral normalization tool that mean centers the spectra and then divide each mean-centered intensity with its standard deviation [32]:

$$I_{\text{new}} = \frac{I_{\text{old}} - I_{\text{mean}}}{\sigma} \tag{1}$$

where I_{new} is the new intensity, I_{old} is the intensity that is being mean centered, I_{mean} is the mean intensity and σ is standard deviation of intensities. Besides these two, MinMax data scaling was also tried. MinMax scaling represents procedure where for each feature, we scale the values according to the formula below, so we have feature values between zero and one:

$$I_{\text{scaled}} = \frac{I - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \tag{2}$$

Proceeding further, we detected and ejected outliers with the help of Isolation Forest algorithm implemented in sci-kit learn. After the outliers have been removed, we fitted Random Forest regressor with aim to find features that give the most contribution to the final result. To achieve this, we actually trained four random forest models, one for each element, to have features that contribute to the each element prediction separately. Feature importances were calculated within random forest algorithm by usage of Gini importance. The higher the value, the more valuable this feature is to the final prediction.

3.2 Hyperparameters tuning and model selection

To find the optimal parameters of the model, we performed GridSearch cross-validation. This validation technique takes the given model parameters and initializes the model of interest with these parameters, splits provided dataset into training and test datasets, fits the model and reports the accuracy of the model through R^2 coefficient. This procedure is done five times in a row for each set of model parameters, where, at the end, for each model algorithm reports the best performance and with which parameters they were obtained. Used metrics to assess the predictive performance of the models were coefficient of determination R^2 and root-mean-square error (RMSE). With optimal parameters

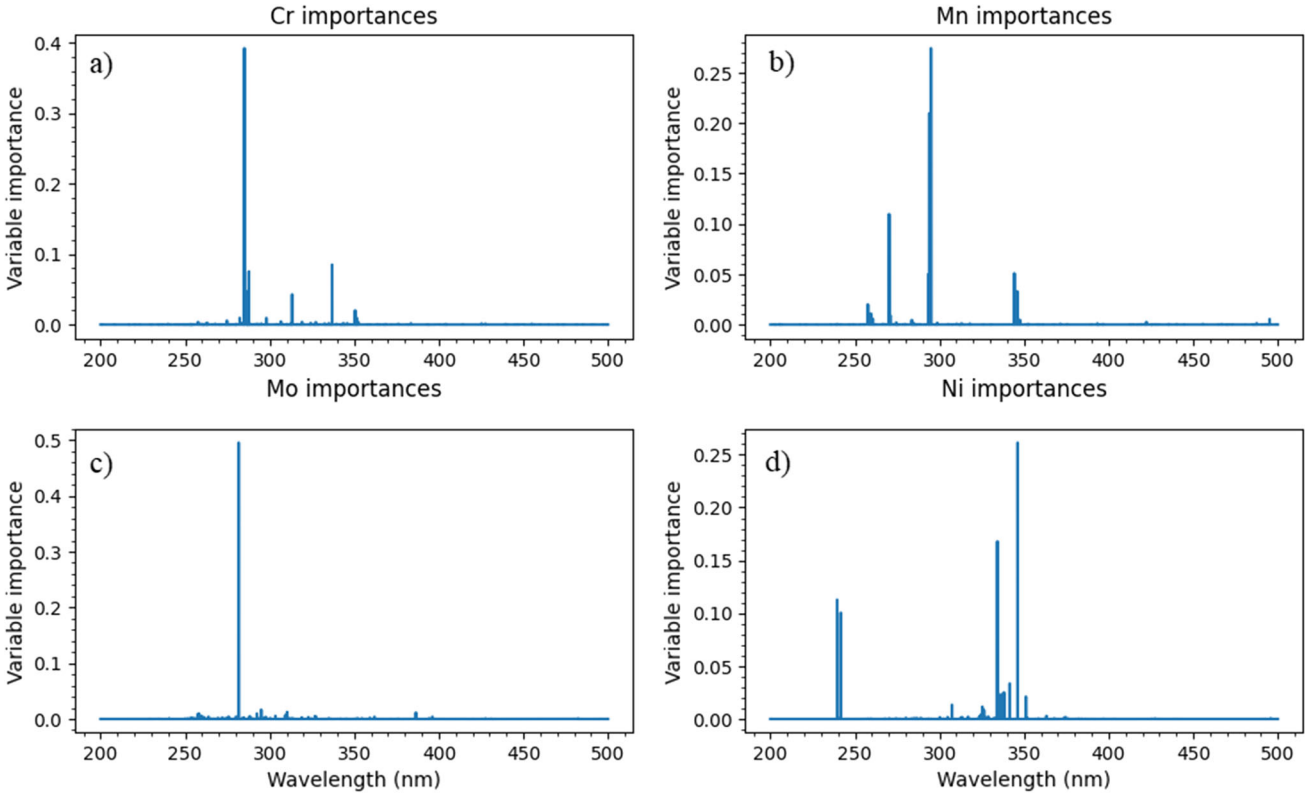


Fig. 3 Results for feature importance analysis for Cr and Mn (**a**, **b**) and Mo and Ni (**c**, **d**)

found, we proceeded to final model training and finally the prediction of steel samples composition.

4 Results

The results of feature importance analysis is given in Fig. 3. It is evident that the algorithm successfully recognized and selected persistent line of Mo II at 281.61 nm (see Fig. 3c)). Also in Fig. 3d), lines of Ni II at 239.45 nm and 241.6 nm were successfully identified. Great importance was also given for Cr II lines around 285, 286, 287, 313 nm, as well as to Cr II line at 336 nm (see Fig. 3a)). Unimportant features have value of zero or close to zero, so the condition threshold was set to $10^{-4}, 2\times10^{-4}$ and 5×10^{-4}, while the best results were obtained for threshold 2×10^{-4}. Hence, the final dimensionality of dataset used to train the final model is given in Table 2.

In GridSearch cross-validation, parameters for random forest that were supplied to the algorithm were number of estimators (number of threes in forest) which was changed from 200 to 350 in the step of 50, and maximal depth of the individual three which was varied from none to 4. None here means that the three is going to expand until all leaves are pure. In the case of linear regression, the only parameter that could be changed is $\mathcal{L}^2$ norm penalization coefficient α, and we have chosen the values of 0.5, 0.8 and 1. For DNN, considered architectures were ones with one, two and three hidden layers [(100), (100, 100) and (100, 150, 50)]. Numbers in parentheses represent number of neurons in each hidden layer. Activation function was ReLU (Rectified Linear Unit). Best results reported for all models were ones with MinMax scaling. For RF, best results were the ones where number of threes was equal to 350 and maximum depth that was set to none. Best results with linear regression were reached for the α parameter equal to 0.8. Finally, for DNN, architecture with three hidden layers showed best performance. After dimensionality reduction via Gini importance, resulting dataset was divided into training and test datasets, keeping 20% of the data for testing. Validation of the models was done by using R^2 metrics and RMSE, and it is given in Table 3.

With model training finished, judging by the R^2 score, best overall performance is showed by deep neural network. The prediction precision for each element goes above 0.9, whereas the predicted values in the case of RF are little less. Results for linear regression are not given, since they are significantly worse than these, thus they were omitted. Prediction on recorded test dataset was done with RF as well as with DNN, and the predicted results are summed in Fig. 4. From Fig. 4a–d, it can be seen that DNN showed good performance on all elements, while the predictions made using RF are quite good for the case of Cr, Mn and Ni, but it showed bad overall performance regarding the prediction of Mo, see Fig. 4d. There was no difference when we tried to

Table 2 Dimensionality and number of samples in training dataset used for model training

Element	Number features	Number of samples
Cr	273	1608
Mn	129	1608
Mo	317	1608
Ni	120	1608

All useful information is contained in these selected features

predict Mo concentration with all features, where unimportant features were not removed.

5 Conclusion and future development

In this paper, the prediction of austenitic steel alloy samples was done using the random forest algorithm and deep neural network. Data preprocessing consisted of applying MinMax scaler on the raw data, followed by outliers removal with isolation forest algorithm. Feature selection was performed by Gini importance criterion within random forest algorithm. It successfully isolated most important features, thus enabling the dimensionality reduction while keeping all the necessary information. This was preceded by final training of three models: random forest, linear regression with $\mathcal{L}^2$ norm and deep neural network. Random forest and neural network showed better predictive power than linear regression; hence, they were used as selected models for prediction of the steel alloy composition. Trained random forest model showed good predictive power for

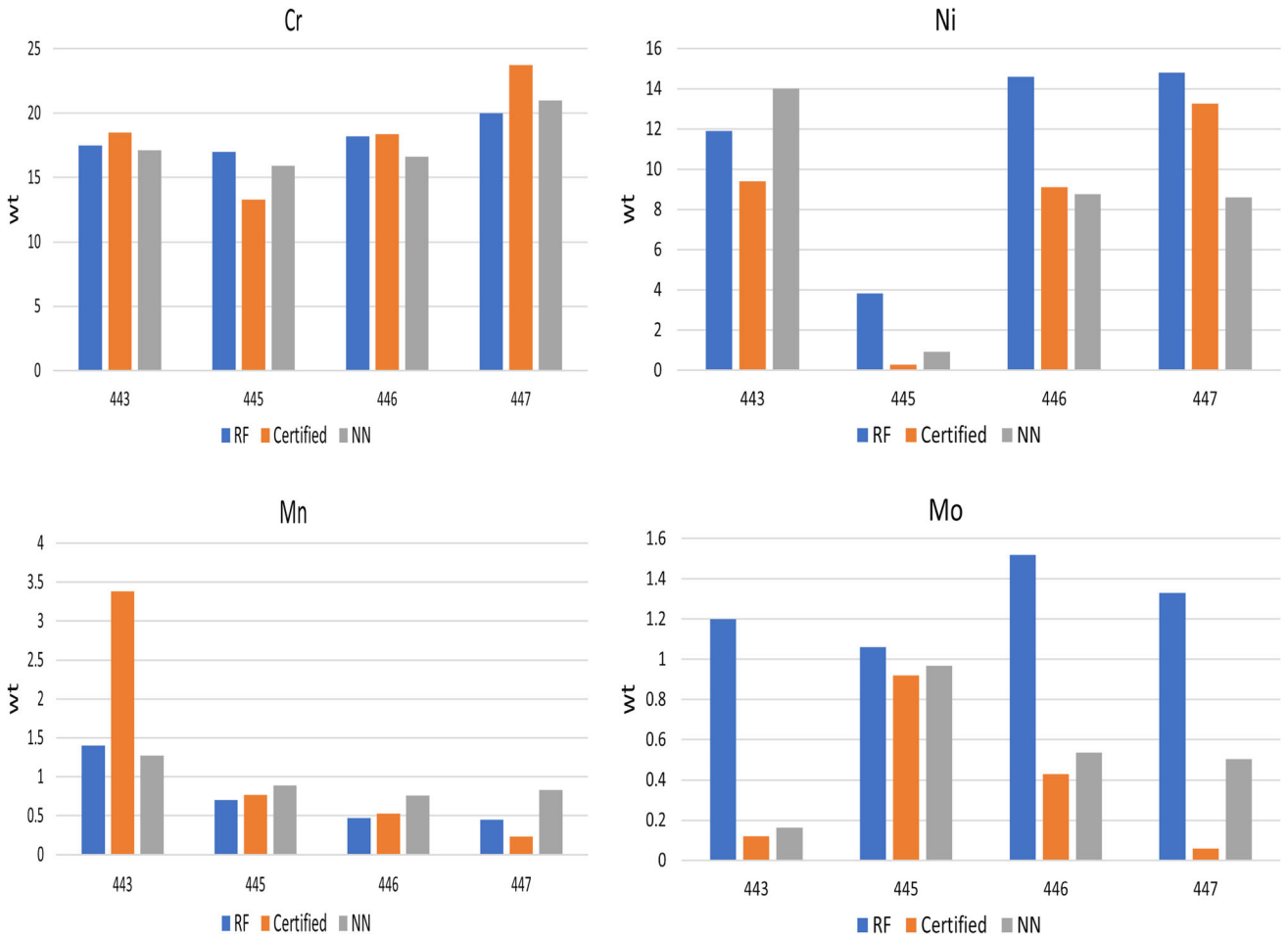


Fig. 4 RFR and DNN predicted results (denoted with RF and NN on the figure) and comparison with certified values. Numbers on x-axis denote the steel sample number given in Table 1. The figure indicates that the models learned and yielded good results in the case of Cr (**a**), Mn (**c**) and Ni (**b**), but on the other hand, RFR had rather poor performance in the case of Mo (**d**)

Table 3 R^2 and RMSE values for validation dataset

Element	R^2_{RFR}	R^2_{DNN}	$RMSE_{RF}$	$RMSE_{NN}$
Cr	0.88	0.97	3.68	1.84
Mn	0.89	0.93	0.397	0.313
Mo	0.85	0.96	0.511	0.263
Ni	0.97	0.98	1.77	1.21

Cr, Ni and Mn, but rather poor performance in the case of Mo. On the other hand, neural network showed good overall predictive power. Nevertheless, random forest algorithm, combined with the data preprocessing techniques, shows a good potential for application in austenitic steel alloy composition prediction, which was also confirmed by results from other authors. For future work, we tend to write a better feature extraction software that should improve the feature selection and hence the predictive power of a used regressor. Also, good overall results are obtained, although the training and test datasets were not intensity corrected. This work shows how useful it can be, to build a unique steel dataset for later usage by different authors, as they not need to every time record their own datasets. These results can be further improved, if one performs calibration transfer, as these spectra were recorded on different instruments. Here, this was not performed as we have not had any identical standard that was used on primary instrument.

Acknowledgements The research was funded by the Ministry of Science, Technological Development and Innovations of the Republic of Serbia, Contract Numbers: 451-03-68/2022-14/200024 and 451-03-68/2022-14/200146, and supported by the Science Fund of the Republic Serbia, Grant No. 3108/2021—NOVA2LIBS4fusion. Also we want to acknowledge the work of our technical associate Stanko Milanović for valuable assistance during the experimental setup. Finally, we want also to thank prof. Jelena Savović for providing us the test samples used in this work.

Author contributions

IT involved in methodology, software and original draft writing; MI involved in conceptualization, supervision and original draft editing.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: As we recorded the spectra of only four steel standard samples, we are of the opinion that scientific community will not have much use of this data, hence it is decided not to deposit it.]

Declarations

Conflict of interest Hereby, we want to state that these results should not be in any case compared to ones obtained by various authors at the LIBS 2022 conference benchmarking competition, where this training dataset was first used. Best results from that competition will be published separately in special issue of Spectrochimica Acta B, and results from this paper have nothing to do with that competition. This paper was an extension of the work presented at SPIG 2022 conference.

References

1. R.J. Konings, R.E. Stoller (eds.), *Comprehensive Nuclear Materials* (Elsevier, Amsterdam, 2020)
2. N. Inoue, A. Komori, H. Hayashi, H. Yonezu, M. Iima, R. Sakamoto, Y. Kubota, A. Sagara, K. Akaishi, N. Noda, N. Ohyabu, O. Motojima, Design and construction of the LHD plasma vacuum vessel. Fus. Eng. Des. **41**(1), 331–336 (1998). https://doi.org/10.1016/S0920-3796(98)00248-8
3. V. Alimov, M. Yajima, S. Masuzaki, M. Tokitani, Analysis of mixed-material layers deposited on the toroidal array probes during the FY 2012 LHD plasma campaign. Fus. Eng. Des. **147**, 111228 (2019). https://doi.org/10.1016/j.fusengdes.2019.06.001
4. D.C.L.J. Radziemski, *Spectrochemical analysis using laser plasma excitation*, edited by D.C.L.J. Radziemski (Marcel Dekker Inc, New York, 1989)
5. E. Tognoni, G. Cristoforetti, S. Legnaioli, V. Palleschi, Calibration-free laser-induced breakdown spectroscopy: state of the art. Spectrochim. Acta Part B At. Spectrosc. **65**(1), 1–14 (2010). https://doi.org/10.1016/j.sab.2009.11.006
6. C.A.M. Ramirez, M. Greenop, L. Ashton, I. ur Rehman, Applications of machine learning in spectroscopy. Appl. Spectrosc. Rev **56**(810), 733–763 (2021). https://doi.org/10.1080/05704928.2020.1859525
7. N.M. Ralbovsky, I.K. Lednev, Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning. Chem. Soc. Rev. **49**, 7428–7453 (2020). https://doi.org/10.1039/D0CS01019G
8. W. Fu, W.S. Hopkins, Applying machine learning to vibrational spectroscopy. J. Phys. Chem. A **122**, 167–171 (2017). https://doi.org/10.1021/acs.jpca.7b10303
9. H. Park, J.-H. Son, Machine learning techniques for THz imaging and time-domain spectroscopy. Sensors (2021). https://doi.org/10.3390/s21041186
10. M. Koubiti, M. Kerebel, Application of deep learning to spectroscopic features of the Balmer-Alpha line for hydrogen isotopic ratio determination in tokamaks. Appl. Sci. (2022). https://doi.org/10.3390/app12199891
11. T. Chen, T. Zhang, H. Li, Applications of laser-induced breakdown spectroscopy (LIBS) combined with machine

learning in geochemical and environmental resources exploration. Trends Anal. Chem. **133**, 116113 (2020). https://doi.org/10.1016/j.trac.2020.116113
12. C. Sun, Y. Tilan, L. Gao et al., Machine learning allows calibration models to predict trace element concentration in soils with generalized LIBS spectra. Sci. Rep. **9**, 11363 (2019). https://doi.org/10.1038/s41598-019-47751-y
13. X. Zhang, F. Zhang, H.-T. Kung, P. Shi, A. Yushanjiang, S. Zhu, Estimation of the Fe and Cu contents of the surface water in the Ebinur lake basin based on LIBS and a machine learning algorithm. Int. J. Environ. Res. Public Health (2018). https://doi.org/10.3390/ijerph15112390
14. L. Sheng, T. Zhang, G. Niu, K. Wang, H. Tang, Y. Duan, H. Li, Classification of iron ores by laser-induced breakdown spectroscopy (LIBS) combined with random forest (RF). J. Anal. At. Spectrom. **30**, 453–458 (2015). https://doi.org/10.1039/C4JA00352G
15. Y. Tian, Q. Chen, Y. Lin, Y. Lu, Y. Li, H. Lin, Quantitative determination of phosphorus in seafood using laser-induced breakdown spectroscopy combined with machine learning. Spectrochim. Acta Part B At. Spectrosc. **175**, 106027 (2021). https://doi.org/10.1016/j.sab.2020.106027
16. M.S. Babu, T. Imai, R. Sarathi, Classification of aged epoxy micro-nanocomposites through PCA- and ANN-adopted LIBS analysis. IEEE Trans. Plasma Sci. **49**(3), 1088–1096 (2021). https://doi.org/10.1109/TPS.2021.3061410
17. X. Cui, Q. Wang, Y. Zhao et al., Laser-induced breakdown spectroscopy (LIBS) for classification of wood species integrated with artificial neural network (ANN). Appl. Phys. B **125**, 12556 (2019). https://doi.org/10.1007/s00340-019-7166-3
18. R. Junjuri, M.K. Gundawar, A low-cost LIBS detection system combined with chemometrics for rapid identification of plastic waste. Waste Manag. **117**, 48–57 (2020). https://doi.org/10.1016/j.wasman.2020.07.046
19. L.-N. Li, X.-F. Liu, F. Yang, W.-M. Xu, J.-Y. Wang, R. Shu, A review of artificial neural network based chemometrics applied in laser-induced breakdown spectroscopy analysis. Spectrochim. Acta Part B At. Spectrosc. **180**, 106183 (2021). https://doi.org/10.1016/j.sab.2021.106183
20. F. Poggialini, B. Campanella, S. Legnaioli, S. Pagnotta, S. Raneri, V. Palleschi, Improvement of the performances of a commercial hand-held laser-induced breakdown spectroscopy instrument for steel analysis using multiple artificial neural networks. Rev. Sci. Instrum. **91**(7), 073111 (2020). https://doi.org/10.1063/5.0012669
21. H. Tang, T. Zhang, X. Yang, H. Li, Classification of different types of slag samples by laser-induced breakdown spectroscopy (LIBS) coupled with random forest based on variable importance (VIRF). J. Anal. At. Spectrom. **32**, 2194–2199 (2017). https://doi.org/10.1039/C7JA00231A
22. F. Ruan, J. Qi, C. Yan, H. Tang, T. Zhang, H. Li, Quantitative detection of harmful elements in alloy steel by LIBS technique and sequential backward selection-random forest (SBS-RF). Anal. Methods **7**, 9171–9176 (2015). https://doi.org/10.1039/C5AY02208H
23. J. Liang, M. Li, Y. Du, C. Yan, Y. Zhang, T. Zhang, X. Zheng, H. Li, Data fusion of laser induced breakdown spectroscopy (LIBS) and infrared spectroscopy (IR) coupled with random forest (RF) for the classification and discrimination of compound *Salvia miltiorrhiza*. Chemom. Intell. Lab. Syst. **207**, 104179 (2020). https://doi.org/10.1016/j.chemolab.2020.104179
24. G. Yang et al., The basicity analysis of sintered ore using laser-induced breakdown spectroscopy (LIBS) combined with random forest regression (RFR). Anal. Methods **9**, 5365–5370 (2017). https://doi.org/10.1039/C7AY01389B
25. L. Breiman, Random forests. Mach. Learn. **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
26. T. Zhang et al., A novel approach for the quantitative analysis of multiple elements in steel based on laser-induced breakdown spectroscopy (LIBS) and random forest regression (RFR). J. Anal. At. Spectrom. **29**, 2323 (2014). https://doi.org/10.1039/c4ja00217b
27. Y. Zhang, C. Sun, L. Gao, Z. Yue, S. Shabbir, W. Xu, M. Wu, J. Yu, Determination of minor metal elements in steel using laser-induced breakdown spectroscopy combined with machine learning algorithms. Spectrochim. Acta Part B At. Spectrosc. **166**, 105802 (2020). https://doi.org/10.1016/j.sab.2020.105802
28. K. Liu et al., Quantitative analysis of toxic elements in polypropylene (PP) via laser-induced breakdown spectroscopy (LIBS) coupled with random forest regression based on variable importance (VI-RFR). Anal. Methods **11**, 4769 (2019). https://doi.org/10.1039/c9ay01796h
29. K. Wei, Q. Wang, G. Teng, X. Xu, Z. Zhao, G. Chen, Application of laser-induced breakdown spectroscopy combined with chemometrics for identification of penicillin manufacturers. Appl. Sci. (2022). https://doi.org/10.3390/app12104981
30. X. Jin, G. Yang, X. Sun, D. Qu, S. Li, G. Chen, C. Li, D. Tian, L. Yao, Discrimination of rocks by laser-induced breakdown spectroscopy combined with random forest (RF). J. Anal. At. Spectrom. **38**, 243–252 (2023). https://doi.org/10.1039/D2JA00290F
31. E. Kepes. (2022) LIBS 2022 quantification contest. https://figshare.com/projects/LIBS2022_Quantification_Contest/142250
32. T.W. Randolph, Scale-based normalization of spectral data. Cancer Biomark. **2**, 135–144 (2006). https://doi.org/10.3233/CBM-2006-23-405

Contrib. Astron. Obs. Skalnaté Pleso **52/3**, 97 – 104, (2022)
https://doi.org/10.31577/caosp.2022.52.3.97

The usage of perceptron, feed and deep feed forward artificial neural networks on the spectroscopy data: astrophysical & fusion plasmas

N.M. Sakan, I. Traparić, V.A. Srećković and M. Ivković

Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia (E-mail: nsakan@ipb.ac.rs)

Received: July 28, 2022; Accepted: October 1, 2022

Abstract. Artificial neural networks are gaining a momentum for solving complex problems in all sorts of data analysis and classification matters. As such, idea of determining their usability on complex plasma came up. The choice for the input data for the analysis is a set of stellar spectral data. It consists of complex composition plasma under vast variety of conditions, dependent on type of star, measured with calibrated standardized procedures and equipment. The results of the analysis has shown that even a simple type of perceptron artificial neural network could lead to results of acceptable quality for the analysis of spectra of complex composition. The analyzed ANNs performed good on a limited data set. The results can be interpreted as a figure of merit for further development of complex neural networks in various applications e.g. in astrophysical and fusion plasmas.

Key words: Atomic processes–Line: profiles-astrophysical & fusion plasmas

1. Introduction

The usage of machine learning algorithms is a growing field of research (D'Isanto et al., 2016; Baron, 2019; Kates-Harbeck et al., 2019). Since the computer power is constantly growing its usage is often found in a wide variety of applications: from determination of objects on a photograph all the way to expert systems capable to determine adequate states and predicted outcomes of complex systems; from difficult-to-maintain machines states and prediction of conditions, up to the assistance in human health monitoring.

Even the specific fields of spectroscopy rely deeply on artificial neural networks, as is the case for instance with medical spectroscopy application Wang et al. (2015), or for instance agricultural application Basile et al. (2022); Longin et al. (2019). The material recognition in extraterrestrial spectroscopic probing is also a very difficult task, since the limitations of the mass and resolving power of the onboard instruments are a very difficult limiting factor (Koujelev et al., 2010; Bornstein et al., 2005).

Usage of the artificial neural networks (ANN) fell into focus of our interest because of flexibility of their application, as well as a variety of complex problems that they have already solved.

All of the mentioned has been a factor for applying neural networks to the decision process of determining a stellar spectral type as an example of application on astrophysical data (Albert et al., 2020). Artificial neural networks are often used in astrophysics (e.g. for the integral field spectral analysis of galaxies in Hampton et al. 2017). There is an expectation of development of further focus on convolutional neural networks application on spectroscopic data (Castorena et al., 2021). Also, even more complex predictions based on back-propagation in neural networks as well as complex artificial neural networks structures in spectroscopic usage are known (Li et al., 2017). In order to have insight of applicability of the ANN usage we have limited our research on simplest case as a figure of merit.

Few random spectral curves from database Pickles (1998) are presented here in results. Entire database set consists of spectra for 12 types of stars, spectral type O normal; B normal; A normal; F normal; F Metal rich; F metal weak; G normal; G Metal rich; G metal weak; K normal; K Metal rich; K metal weak; and M normal. Our aim was to create test case as a method of determining a quality of specific ANN in various machine learning analysis, from stellar and fusion spectra analysis, material analysis, up to extremely specific cases as enhancing a low resolution instrument performance for specific applications (Marinković et al., 2019; Albert et al., 2020).

2. ANN basics and principles

The usage of systems related to the functions of neural networks has been in focus of investigation since mid-1940 McCulloch & Pitts (1943), but the real usage has evolved with the application of modern day digital computers, which enabled construction of networks of enlarged complexity. One of the simplest neural networks, that could be seen more as a test case of validity of operation of artificial intelligence systems, is perceptron (Rosenblatt, 1958). The prediction as well as sensitivity of the training data set is in favor of more complex networks. It is the primary goal of our investigation, along with their application on spectral data sets and measurements.

The choice for the dataset was made on open access data files for the 131 stellar spectra published by Pickles (1998) (available at accompanying reference appended to the bibliographical entry, as seen in May 2022). The results are promising and further research on the field is expected. The quality of the trained artificial neural network prediction is related to the data set as well as its structure. An effort of applying it on a large scale dataset or database should be carried out.

The problem of finding out a category of data subset is an inherent problem for any sort of machine learning and as such for the artificial neural networks also. The artificial neural network is a system of mathematical functions trying to resemble a simplified animal brain. The network consists of artificial neurons.

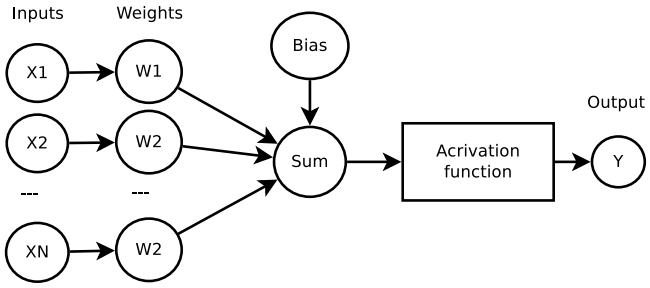


Figure 1. The concept of a neuron. Schematic presentation.

A neuron is described as a function that adopts output value based on its input values and bias value by the means of reaction function. The simplest neuron concept could be seen on a Figure 1. The neuron determines its output state as an output of activation function based on a weighted sum of input values and a bias value itself, and could be described by equation

$$y_{out} = f_{act}\left(Bias + \sum_{i=1}^{N} x_i w_i\right), \tag{1}$$

where f_{act} is a activation function, x_i and w_i are the i-th input value as well as adequate input weight.

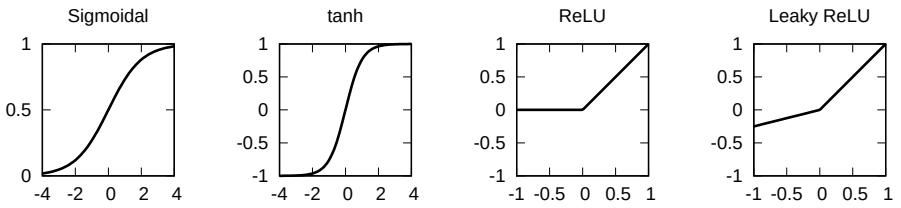


Figure 2. Four most common neuron reaction functions.

The neuron reaction on external stimulus is strongly dependent on its reaction function. In order to determine the neuron behavior on a micro scale, the reaction function as well as the method of adopting the weight values plays a determining role. Four most common reaction functions are shown on Figure 2.

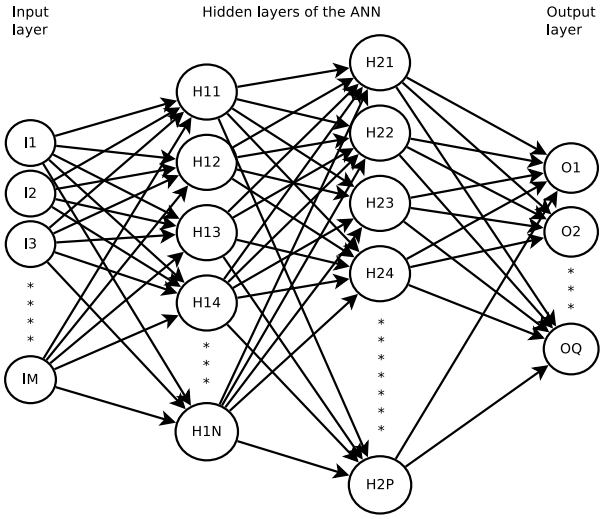


Figure 3. Concept of ANN of perceptron, feed forward and deep feed forward.The shown ANN consist of M input neurons, two hidden layers of N and P neurons and output layer of Q neurons.

A topology of the neural network as well as the learning method are determining the global reaction of the neural network. For the goal of usability analysis the simplest ANN topologies, perceptron, is chosen. The Feed Forward and Deep Feed Forward topologies are based upon fully connected dense layers of neurons, see Figure 3. The two specific layers, input and output, have the dimensionality of the input data and output states consequently and are the only limiting factors of the network. When there is more than one hidden layer, the neural network is considered to be the deep one.

3. Results and discussion

In Figure 4 several random spectral curves from database Pickles (1998) are presented. Entire dataset consists of spectra for 12 types of stars, spectral type O normal; B normal; A normal; F normal; F Metal rich; F metal weak; G normal; G Metal rich; G metal weak; K normal; K Metal rich; K metal weak; and M normal. Each epoch of the dataset was divided into 70% for training set and 30% for the test set.

As a test bench for the application of the ANN to the selection set of perceptron, Feed Forward and Deep Feed Forward networks are used. As a reaction function ReLU (rectified linear unit) was used, and the input data was normalized to unit using standardization

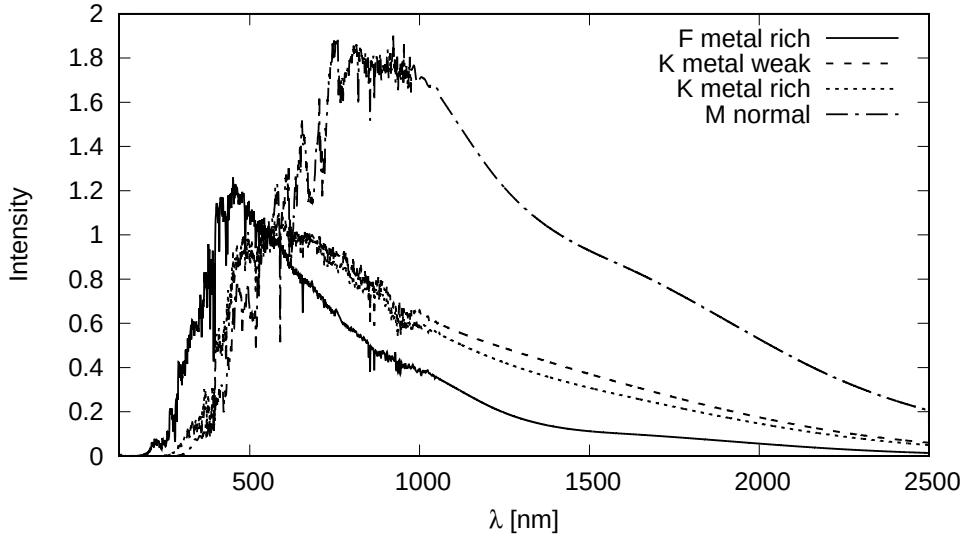


Figure 4. Several sample spectra. Spectra i.e. data is taken from Pickles (1998).

$$x' = \frac{x - \mu}{\sigma}, \qquad \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad \hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{\mu})^2}. \tag{2}$$

No additional data preparation was imposed. The investigated neural network topologies were let to train on the set of data for 200 epochs. Input layer consisted of 4771 input values of available data, output layer consisted of 12 types of stars, spectral type O normal; B normal; A normal; F normal; F Metal rich; F metal weak; G normal; G Metal rich; G metal weak; K normal; K Metal rich; K metal weak; and M normal. The hidden layers consisted of 5000 neurons in first, 1000 in second and 512 neurons in third layer. They were included consequently in order to compare ANN behavior, see Figure 5.

It is obvious, by the analysis of calculated data presented in Figure 5, that the deeper ANNs are capable to learn faster and have better predictions after smaller epochs of learning. This capability is a winning solution in the case of complex spectra. The ANN could fall into pseudo stable states and produce a non-minimal error. Such falls into local minimum state could be avoided by several advanced methods one of which is providing an algorithm for forgetting of the learned state, e.g. algorithm that disturbs a learned state after each application.

Also, there are probably better methods for the input dataset preparation, from pure mathematical procedures up to convolutional ANN (CNN) incorporation. It is proven that, even in its simplest forms, ANN could be used for such

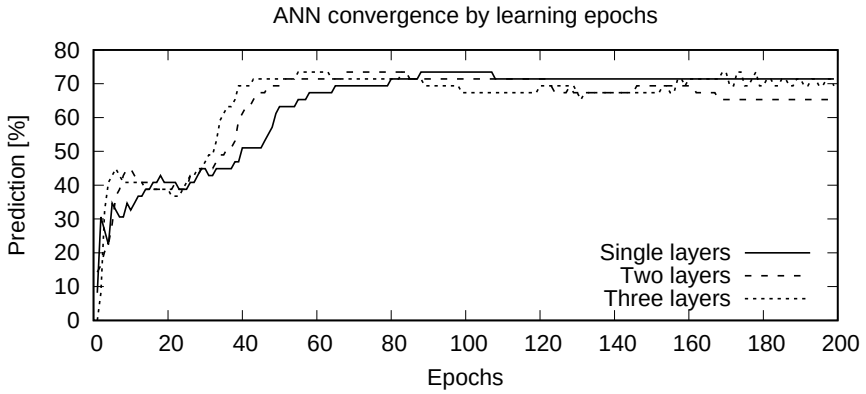


Figure 5. Convergence of ANN of perceptron, feed forward and deep feed forward, on analyzed dataset.

tasks. It is to expect that there are better ANN topologies for such a task, and this is a field for further investigation.

From the above it is obvious that even in its crudest form artificial neural networks are capable to successfully deal with the spectra classification. It is confirmed that this case could be used as a figure of merit for the further development of ANN and machine learning applications in general.

4. Conclusions and future possibilities

The results are promising and the further research on the field is expected. The first goal of analysis of a single set of complex spectral data recorded under similar conditions is achieved with reasonably good prediction. Concerning minute differences in comparison to each other it is considerable result for the basic ANN structure.

Since the quality of the trained artificial neural network prediction is related to its structure as well as the dataset quality and volume, an effort on a large-scale database collection should be carried out. One of the first steps should be inclusion of pre-trained convolutional ANN for the purpose of input data pre-processing before entering of selector ANN.

Commercial packages as well as some specific open-source solutions for the analysis of the spectra with the help of predefined ANN exist. Their application is usually very specific and does not allow the opportunity to fit the best ANN nor to perform unique mathematical procedures during input data preparation that could be best suited for the sought purpose. This possibility is the winning factor in each specific case. Such approach should enable systems for more specialized problem solutions, from stellar and fusion spectra analysis up to more

specific expert systems related to technical solutions. The further development in both ANN structures as well as data preparation should be carried out with the specified problem in mind.

Acknowledgements. The research was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Contract: 451-03-68/2022-14/200024 and supported by the Science Fund of the Republic Serbia, Grant no. 3108/2021.

References

Albert, D., Antony, B. K., Ba, Y. A., et al., A Decade with VAMDC: Results and Ambitions. 2020, *Atoms*, **8**, 76, DOI: 10.3390/atoms8040076

Baron, D., Machine Learning in Astronomy: a practical overview. 2019, *arXiv e-prints*, arXiv:1904.07248

Basile, T., Marsico, A. D., & Perniola, R., Use of Artificial Neural Networks and NIR Spectroscopy for Non-Destructive Grape Texture Prediction. 2022, *Foods*, **11**, DOI: 10.3390/foods11030281

Bornstein, B., Castano, R., Gilmore, M., Merrill, M., & Greenwood, J., Creation and testing of an artificial neural network based carbonate detector for Mars rovers. 2005, in *2005 IEEE Aerospace Conference*, 378–384

Castorena, J., Oyen, D., Ollila, A., Legett, C., & Lanza, N., Deep spectral CNN for laser induced breakdown spectroscopy. 2021, *Spectrochimica Acta Part B: Atomic Spectroscopy*, **178**, 106125, DOI: https://doi.org/10.1016/j.sab.2021.106125

D'Isanto, A., Cavuoti, S., Brescia, M., et al., An analysis of feature relevance in the classification of astronomical transients with machine learning methods. 2016, *Monthly Notices of the RAS*, **457**, 3119, DOI: 10.1093/mnras/stw157

Hampton, E. J., Medling, A. M., Groves, B., et al., Using an artificial neural network to classify multicomponent emission lines with integral field spectroscopy from SAMI and S7. 2017, *Monthly Notices of the Royal Astronomical Society*, **470**, 3395, DOI: 10.1093/mnras/stx1413

Kates-Harbeck, J., Svyatkovskiy, A., & Tang, W., Predicting disruptive instabilities in controlled fusion plasmas through deep learning. 2019, *Nature*, **568**, 526, DOI: 10.1038/s41586-019-1116-4

Koujelev, A., Sabsabi, M., Motto-Ros, V., Laville, S., & Lui, S., Laser-induced breakdown spectroscopy with artificial neural network processing for material identification. 2010, *Planetary and Space Science*, **58**, 682, DOI: https://doi.org/10.1016/j.pss.2009.06.022, exploring other worlds by exploring our own: The role of terrestrial analogue studies in planetary exploration

Li, Z., Zhang, X., Mohua, G. A., & Karanassios, V., Artificial Neural Networks (ANNs) for Spectral Interference Correction Using a Large-Size Spectrometer and ANN-Based Deep Learning for a Miniature One. 2017, in *Advanced Applications for Artificial Neural Networks*, ed. A. El-Shahat (Rijeka: IntechOpen)

Longin, L., Tusek, A. J., Valinger, D., et al., Application of Artificial Neural Networks (ANN)Coupled with Near-InfraRed(NIR) Spectroscopy for Detection of Adulteration in Honey. 2019, *Biodiversity Information Science and Standards*, **3**, e38048, DOI: 10.3897/biss.3.38048

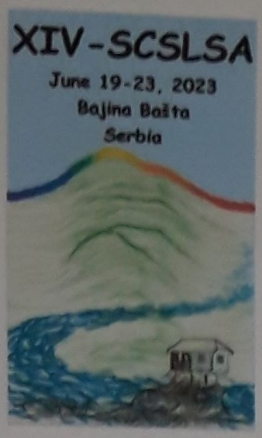
Marinković, B., Srećković, V., Vujčić, V., et al., BEAMDB and MOLD—Databases at the Serbian Virtual Observatory for Collisional and Radiative Processes. 2019, *Atoms*, **7**, 11, DOI: 10.3390/atoms7010011

McCulloch, W. S. & Pitts, W., A logical calculus of the ideas immanent in nervous activity. 1943, *The bulletin of mathematical biophysics*, **5**, 115, DOI: 10.1007/BF02478259

Pickles, A. J., A Stellar Spectral Flux Library: 1150-25000 Å. 1998, *Publications of the ASP*, **110**, 863, DOI: 10.1086/316197

Rosenblatt, F., The perceptron: A probabilistic model for information storage and organization in the brain. 1958, *Psychological Review*, **65**, 386, DOI: 10.1037/h0042519

Wang, Q., Zheng, N., Li, Z., & Ma, Z., Quantitative analysis of glucose in whole blood using FT-Raman spectroscopy and artificial neural network. 2015, in *Proceedings of the 2015 International Conference on Computational Science and Engineering* (Atlantis Press), 471–475



Ivan Traparić
Institute of Physics,
Belgrade, Serbia

14th Serbian Conference on Spectral Line Shapes in Astrophysics
Bajina Bašta, June 19-23, 2023, Serbia

23.06.2023.

We certify that Ivan Traparić has presented a progress report "Strak Broadening Modeling with ML and AI algorithms" in the *14th Serbian Conference on Spectral Line Shapes in Astrophysics* (14th SCSLSA), held in 19-23. June 2023. in Bajina Bašta, Serbia.

Sincerely,

Luka. Č. Popović
(Chairman of the Scientific Committee)

Nataša Bon
(Chairman of the Local Organizing Committee)

Република Србија
Универзитет у Београду
Физички факултет
Д.Бр.2020/8009
Датум: 08.08.2023. године

На основу члана 161 Закона о општем управном поступку и службене евиденције издаје се

УВЕРЕЊЕ

Трапарић (Радомир) Иван, бр. индекса 2020/8009, рођен 14.09.1996. године, Требиње, Босна и Херцеговина, уписан школске 2022/2023. године, у статусу: финансирање из буџета; тип студија: докторске академске студије; студијски програм: Физика.

Према Статуту факултета студије трају (број година): три.
Рок за завршетак студија: у двоструком трајању студија.

Ово се уверење може употребити за регулисање војне обавезе, издавање визе, права на дечији додатак, породичне пензије, инвалидског додатка, добијања здравствене књижице, легитимације за повлашћену вожњу и стипендије.

Овлашћено лице факултета

Република Србија
Универзитет у Београду
Физички факултет
Д.Бр.2020/8009
Датум: 08.08.2023. године

На основу члана 161 Закона о општем управном поступку и службене евиденције издаје се

УВЕРЕЊЕ

Трапарић (Радомир) Иван, бр. индекса 2020/8009, рођен 14.09.1996. године, Требиње, Босна и Херцеговина, уписан школске 2022/2023. године, у статусу: финансирање из буџета; тип студија: докторске академске студије; студијски програм: Физика.

Према Статуту факултета студије трају (број година): три.
Рок за завршетак студија: у двоструком трајању студија.

Ово се уверење може употребити за регулисање војне обавезе, издавање визе, права на дечији додатак, породичне пензије, инвалидског додатка, добијања здравствене књижице, легитимације за повлашћену вожњу и стипендије.

Овлашћено лице факултета

Република Србија

Универзитет у Београду

Оснивач: Република Србија

Дозволу за рад број 612-00-02666/2010-04 од 12. октобра 2011. године је издало Министарство просвете и науке Републике Србије

УБ

Физички факултет, Београд

Оснивач: Република Србија

Дозволу за рад број 612-00-02409/2014-04 од 8. септембра 2014. године је издало Министарство просвете, науке и технолошког развоја Републике Србије

Диплома

Иван, Радомир, Трајарић

рођен 14. септембра 1996. године, Требиње, Босна и Херцеговина, уписан школске 2015/2016. године, а дана 27. септембра 2019. године завршио је основне академске студије, првог степена, на студијском програму Примењена и комјутерска физика, обима 240 (двеста четрдесет) бодова ЕСПБ са просечном оценом 9,43 (девет и 43/100).

На основу тога издаје му се ова диплома о стеченом високом образовању и стручном називу

дипломирани физичар

Број: 11280500

У Београду, 29. октобра 2020. године

Декан
Проф. др Иван Белча

Ректор
Проф. др Ивanka Поповић

00112929

Република Србија

Универзитет у Београду

Оснивач: Република Србија

Дозволу за рад број 612-00-02666/2010-04 од 12. октобра 2011. године је издало Министарство просвете и науке Републике Србије

УБ

Физички факултет, Београд

Оснивач: Република Србија

Дозволу за рад број 612-00-02409/2014-04 од 8. септембра 2014. године је издало Министарство просвете, науке и технолошког развоја Републике Србије

Диплома

Иван, Радомир, Трајарић

рођен 14. септембра 1996. године, Требиње, Босна и Херцеговина, уписан школске 2019/2020. године, а дана 29. септембра 2020. године завршио је мастер академске студије, другог степена, на студијском програму Теоријска и експериментална физика, обима 60 (шездесет) бодова ЕСПБ са просечном оценом 10,00 (десет и 0/100).

На основу тога издаје му се ова диплома о стеченом високом образовању и академском називу

мастер физичар

Број: 11279100

У Београду, 29. октобра 2020. године

Декан
Проф. др Иван Белча

Ректор
Проф. др Иванка Поповић

00112943

	ДОКТОРСКЕ СТУДИЈЕ		

ПРЕДЛОГ ТЕМЕ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ КОЛЕГИЈУМУ ДОКТОРСКИХ СТУДИЈА

Школска година 2022/2023

Подаци о студенту

Име	Иван
Презиме	Трапарић
Број индекса	8009/2020

Научна област дисертације: Физика јонизованог гаса и плазме

Подаци о ментору докторске дисертације

Име	Маријана
Презиме	Гавриловић Божовић
Научна област	Електротехника и рачунарство
Звање	доцент
Институција	Факултет инжењерских наука, Универзитет у Крагујевцу

Предлог теме докторске дисертације

Наслов

Примена вештачке интелигенције и машинског учења у спектроскопији плазме

Уз пријаву теме докторске дисертације Колегијуму докторских студија, потребно је приложити следећа документа:

1. Семинарски рад (дужине до 10 страница)
2. Кратку стручну биографију писану у трећем лицу једнине
3. Фотокопију индекса са докторских студија

Потпис ментора

Датум 11.4.2023.

Потпис студента

Мишљење Колегијума докторских студија

Након образложења теме докторске дисертације Колегијум докторских студија је тему

прихватио ☒ није прихватио ☐

Датум

05.07.2023

Продекан за науку Физичког факултета