Cu fractionation, isotopic analysis, and data processing *via* machine learning: new approaches for the diagnosis and follow up of Wilson's Disease *via* ICP-MS

M. Carmen García-Poyo,[a,b] Sylvain Bérail,[a] Anne Laure Ronzani,[a] Luis Rello,[c] Elena García-González,[c] Flávio V. Nakadi,[b] Maite Aramendía,[b,d] Javier Resano,[e] Martín Resano[b*] and Christophe Pécheyran[a*]

[a] Institut des Sciences Analytiques et de Physico-chimie pour l'Environnement et les Matériaux, UPPA/CNRS 5254, Université de Pau et des Pays de l'Adour, 2 Av Président Angot, Pau, 64000, France, E-mail: christophe.pecheyran@univ-pau.fr

[b] Department of Analytical Chemistry, Aragon Institute of Engineering Research (I3A), University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain, E-mail: mresano@unizar.es

[c] Department of Clinical Biochemistry, IIS Aragón, "Miguel Servet" University Hospital, Paseo Isabel La Católica 1-3, 50009 Zaragoza, Spain

[d] Centro Universitario de la Defensa de Zaragoza, Carretera de Huesca s/n, 50090, Zaragoza, Spain

[e] Department of Computer Sciences and Systems Engineering (DIIS), Aragon Institute of Engineering Research (I3A), University of Zaragoza, C/ Mariano Esquillor SN, 50018, Zaragoza, Spain

Abstract

Information about Cu fractionation and Cu isotopic composition can be paramount when investigating Wilson's disease (WD). This information can provide a better understanding of the metabolism of Cu. Most important, it may provide an easy way to

diagnose and to follow the evolution of WD patients. For such purpose, protocols for Cu determination and Cu isotopic analysis *via* inductively coupled plasma mass spectrometry were investigated in this work, both in bulk serum and in the exchangeable copper (CuEXC) fractions.

The CuEXC protocol provided satisfactory recovery values. Also, no significant mass fractionation during the whole analytical procedure (CuEXC production and/or Cu isolation) was detected. Analyses were carried out in controls (healthy persons), newborns, patients with hepatic disorders, and WD patients. While results for Cu isotopic analysis show relevant results (*e.g.*, $\delta^{65}Cu$ were lower for both WD patients under chelating treatment and patients with hepatic problems in comparison with those values obtained for WD patients under Zn treatments, controls, and newborns) to comprehend Cu metabolism and to follow up the disease, the parameter that can help to better discern between WD patients and the rest of the patients tested (non-WD) was found to be the REC (relative exchangeable Cu).

In this study, all the WD patients showed a REC higher than 17 %, while the rest showed lower values. However, since establishing a universal threshold is complicated, machine learning was investigated to produce a model that can differentiate among WD and non-WD samples with excellent results (100 % accuracy, albeit for a limited sample set). Most important, unlike other ML approaches, our model can also provide an uncertainty metric to indicate the reliability of the prediction, overall opening new ways to diagnose WD.

1. Introduction

Wilson's disease (WD) is a recessive genetic disease that affects 1 out of 30000 people. It is produced by a mutation in the ATP7B gene encoding a transmembrane ATPase protein, which is the central regulator of hepatic copper. This protein plays a role as a Cu ATPase transporter, especially in the liver. It is involved in both the excretion of Cu from the hepatocytes into the bile and the incorporation of Cu into the ceruloplasmin.[1–3] In WD patients, a defective ATP7B prevents the body to remove the excess of copper. As a result, Cu levels build up in various organs, mostly liver and brain, but also kidneys and eyes (Kayser-Fleischer rings). WD is an irreversible disease, which causes serious clinical injuries and even death, if it is misdiagnosed or the treatment is delayed.[3,4] In general, the survival rate depends on the stage of liver and/or neurological disease, and also on treatment follow-up. At present, there are two kinds of treatments for WD: one is based on the administration of chelators for increasing excretion of the Cu that is already accumulated in the organs, such as D-penicillamine and Trientine; the other is based on blocking the absorption of Cu by treatment with Zn salts.[5,6] The choice of treatment depends on the clinical picture of the patient, but both of them are easy to apply.

Therefore, the main challenge associated with WD is to be able to properly diagnose this disease before the onset of symptoms, to avoid the consequences that they entail. Diagnosis can be based on such symptoms, such as detection of Kayser-Fleischer rings in the eyes, liver damage and/or neurological disorders, which is not ideal, or be based on family screening in the case of asymptomatic patients. Genetic tests could also be useful to diagnose this disease. However, they cannot be routinely undertaken because they are costly and tedious, as there are more than 500 genetic mutations that cause WD.[7–9] For this reason, the diagnosis mostly relies on such clinical

symptoms and/or on biological tests, such as blood or urine analyses: *e.g.* total blood/serum Cu, blood/serum ceruloplasmin levels or urine Cu excretion in 24 h, among others.[10] Unfortunately, these kinds of tests may provide unreliable and/or inconclusive results.

For instance, if we focus on serum Cu concentration, this value ranges between 0.7 to 1.5 mg L^{-1} for healthy people,[11] while the typical values found in WD are lower; however, similar values have been found for some controls, particularly for newborns, due to the immaturity of their liver. Values in that range (0.932-0.999 mg L^{-1}) have also been reported for some WD patients, due to the eventual release of Cu from the liver.[12]

In the case of 24 h urinary Cu excretion, the cutoff is set at an excreted Cu amount higher than 100 µg in 24 h. However, asymptomatic WD patients sometimes show an excreted Cu amount lower than such threshold (57-95 µg). These are only some examples of the issues preventing efficient WD diagnosis. As a consequence, it is often necessary to carry out several tests to confirm the disease and, even then, sometimes the diagnosis is not clear.[3,13,14]

Hence the importance of continuing the research on new methods to diagnose WD at an early stage before any symptoms appear. In this regard, investigations focused on Cu isotopic analysis[12,15,16] or on parameters related with the Cu bound to ceruloplasmin and/ or free Cu[13,17-19] have been carrying out.

Isotopic analysis has demonstrated its potential for biomedical investigations,[20-23] particularly when using multicollector inductively coupled plasma mass spectrometry (MC-ICP-MS).[24] In the case of WD, preliminary studies showed that samples from WD patients present a lighter Cu isotopic composition than those obtained from controls, both in serum and in urine samples.[15,16] These results have been further supported by

4

recent work, where it was observed that such lighter isotopic compositions seem to be linked to Cu release from the liver, regardless of whether this release is due to the medical treatment with a chelating agent or due to liver damage (cirrhosis).[12]

On the other hand, besides these studies focused on isotopic analysis, a few works have proposed simple approaches for Cu fractionation, differentiating between the amount of Cu that is either strongly bound to some compounds or more freely available. This approach *a priori* seems interesting in this context, considering that WD affects the binding of Cu to ceruloplasmin. Concepts such as ultrafiltrable Cu (CuUF), which is the Cu bound to low molar mass molecules (amino acids), and exchangeable Cu (CuEXC), which corresponds to the Cu fraction that is not bound to ceruloplasmin and is easily complexed with high-Cu-affinity chelating agents such as ethylenediaminetetraacetic acid (EDTA), have been defined.[25,26] The calculation of relative exchangeable Cu (REC) has been proposed as follows:

$$REC = \frac{CuEXC}{Total\,Cu} \times 100 \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(1)$$

El Balkhi *et al.* described a method to obtain CuUF and CuEXC and established reference values (4.5-9.9 and 36-71 µg L^{-1}, respectively) for healthy individuals.[27] While CuUF was found to be hardly relevant for WD diagnosis, CuEXC and REC were proposed as new biological markers for WD diagnosis.[18] In studies with animal models, Schmiltt *et al.*[17] observed that CuEXC values were correlated with acute liver disease and with the dietary Cu intake. Guillaud *et al.*[19] observed for humans that CuEXC levels were higher for WD patients (symptomatic when diagnosed) and for patients that did not follow the treatment correctly with respect to those WD patients that followed the treatments properly or were asymptomatic; and also, CuEXC values were high for patients with other liver disorders, depending on the stage of the disorder. Overall, the

results reported indicate that CuEXC is not specific for WD, but it could be useful also in this context.

REC, on the other hand, seems to be a sensitive (which in this context means showing a high true positive rate) and specific (meaning showing a high true negative rate) biomarker for WD diagnostics, providing values close to 100 % in terms of sensitivity and specificity. However, a review of the literature reveals that different cutoffs have been proposed by the different authors.[17-19,27,28] Hence, it is still necessary to continue this line of research in order to clarify this critical parameter.

It has to be also mentioned that fractionation approaches based on ultrafiltration are very simple ones, and that other articles using liquid chromatography coupled to inductively coupled plasma mass spectrometry (ICP-MS) have revealed the more complex nature of the different Cu species present in serum[29-31] and the need to keep under strict maintenance the ultrafiltration conditions to keep the specificity of the CuEXC value.[30]

This line of work (species fractionation) can also be further reinforced with isotopic analysis, in order to ascertain if some particular fraction contains more specific information to diagnose and follow up the disease. Lauwens *et al.* used this approach determining bulk, exchangeable and ultrafiltrable serum copper in healthy and alcoholic cirrhosis subjects.[32] However, to the best of the authors knowledge, such strategy has not been investigated any further, neither for WD nor for any other clinical condition.

Besides the need for a more significant number of results, which is always difficult when investigating rare diseases, the use of more advanced chemometrical approaches, such as machine learning (ML),[33] combining the results for the different

Cu species may help in improving the reliability of the diagnosis. It has already been demonstrated that, in complex scenarios where different parameters are monitored, applying the latest ML strategies can help in providing a more accurate classification, although the number of studies reporting on the use of such an approach using trace elemental information of clinical samples is still very limited.[34] Even if ideally for building a reliable model it would be better to get access to information from a number of samples as large as possible, it has been recently reported in a similar situation (the use Cu concentration and isotopic ratios to build a ML model to enhance diagnosis for bladder cancer)[33] that a valuable model can be built even from a limited set of samples.

The aim of the present work is to better understand WD and to try to diagnose this disease more efficiently. For this purpose, the two types of diagnostic approaches for WD (Cu fractionation and Cu isotope analysis) will be explored and combined. For this, the Cu concentration and the isotopic composition in both bulk serum (total Cu) and exchangeable Cu fractions from a series of healthy people (controls), patients with liver diseases different from WD, newborns, and WD patients under different treatments (chelators and/or Zn salts) were determined *via* ICP-MS. Sample ultrafiltration and simple direct microinjection of the samples was preferred over more complex analytical methodologies explored before,[12,35] in order to produce protocols that could be replicated in bio analytical labs, while maintaining the minimal sample consumption needed to carry out all the intended analyses.

Moreover, ML has also been applied to these results for the first time, in order to help in establishing a more efficient way to diagnose WD. We have used ML with two objectives. On the one hand, we want to verify if the data collected contains sufficiently clear and relevant information for the ML models in order to achieve accurate WD diagnosis. On the other hand, we want to enrich the ML output with a metric that reports

the uncertainty of the predictions. The importance both aspects in the context of potentially automated medical diagnosis will be demonstrated.

2. Experimental

2.1. Instrumentation

The development and evaluation of the CuEXC method was carried out with an ELAN DRC II quadrupole ICP-MS and with a NexION 300X ICP-MS, both from Perkin Elmer (Waltham, USA). The conventional configuration using a peristaltic pump to deliver the sample to the nebulizer was chosen as the sample introduction system. Conditions are shown in **Table 1**.

The determinations of total Cu and CuEXC in the samples were carried out with an ELAN DRC II quadrupole ICP-MS (Perkin Elmer). A PFA-ST MicroFlow Nebulizer (Elemental Scientific, Nebraska, USA) coupled to a twister cyclonic spray chamber (Glass expansion, Port Melbourne, Australia) was used for this purpose. The instrumental conditions are shown in **Table 1**.

Cu isotopic analyses were carried out using a Nu 1700 MC-ICP-MS instrument (Nu Instruments, Wrexham, UK), coupled to an Aridus3 (Cetac Teledyne, USA) as desolvating system, which includes a MicroMist 100 µL min^{-1} nebulizer. The instrumental parameters are shown in **Table 2**.

For sample digestion, an ULTRAWAVE microwave system (Milestone Inc., Shelton, USA) was used. An EVAPOCLEAN® unit (Analab, Bischheim, France) was deployed to evaporate samples by sub-boiling in Teflon vessels in a closed environment. A Centrifuge 5810R (Eppendorf, Hamburg, Germany) was used for CuEXC separation.

2.2. Materials and reagents

For CuEXC separation, Amicon® Ultra-4 centrifugal filter devices with 30-kDa cut-off Ultracel® regenerated cellulose membrane were acquired from Merck Millipore (Cork, Ireland). EDTA purchased from Sigma-Aldrich (St. Louis, USA) and potassium chloride from Merck (Darmstadt, Germany) were also used for this purpose.

Poly-prep polypropylene chromatography columns (Bio-Rad, Temse, Belgium) and Cu-specific resin (Triskem, Bruz, France) were used for Cu isolation prior to isotopic analyses. For the preparation of the different eluting solutions, HCl (w = 35 %) ultratrace® (Scharlau, Barcelona, Spain) was used.

For quantitative analysis, single-element standard solutions of Cu and Ni (1 g L^{-1}), were acquired from SCP SCIENCE (Villebon-Sur-Yvette, France). For Cu isotopic analysis, NIST SRM 3114 (NIST, Gaithersburg, USA) was chosen as a reference in the bracketing sequence. This standard shows a similar composition to the NIST SRM 976, which is currently out of stock.[36] NIST SRM 986 (Ni) was used as internal standard for mass bias correction.

To validate the CuEXC protocol, Seronom Trace Elements Serum level 1 (L-1; Lot: 1309438) and 2 (L-2; Lot: 1309416) (Sero, Billingstad, Norway) were used. These reference materials (RM) were reconstituted with 3 mL of ultrapure water, according to the manufacturer's instructions.

Instra grade HNO$_3$ (w = 70 %) was purchased from JT Baker (New Jersey, USA) and further purified by sub-boiling in a PFA system (DST 1000, Savillex, Eden Prairie, USA). Ultrapure water (18.2 MΩ cm) was obtained from a Direct-Q3 system (Millipore, Mollsheim, France).

2.3. Samples and sample preparation

Serum samples from 56 volunteers were obtained from the Hospital Miguel Servet (Zaragoza, Spain). Thirteen of these samples were controls, fourteen originated from patients with hepatic disorders (different from WD), eight came from newborns/infants and twenty-one samples originated from WD patients. The principles outlined in the declaration of Helsinki regarding all the experimental research involving humans or animals were followed. Experiments were approved by the Clinical Research Ethics Committee of Aragon (CEICA). Informed consents were obtained from human participants of this study.

To avoid sample contamination, sample preparation was carried out in an ISO 5 laminar bench flow fitted in an ISO 7 clean lab. Each serum sample was split into two fractions, one for total Cu and the other for CuEXC determination. For each fraction, both Cu determination and Cu isotopic analysis were carried out. This protocol was first validated with serum reference materials (L-1 and L-2), and then applied to real samples.

The first step was to obtain the CuEXC fraction. The procedure proposed by El Balkhi et al.[29] was followed but, instead of using NaCl, KCl was selected to avoid the introduction of additional sources of Na, thus preventing spectral overlap ($^{40}Ar^{23}Na^+$) for the monitoring of $^{63}Cu^+$ in the ICP-MS. 200 µL of serum were mixed in an Eppendorf tube with 200 µL of 3 g L^{-1} EDTA in 9 g L^{-1} KCl. The tube was then vortexed for 20 s. After that, the mixture was incubated for at least 1 h to ensure that the equilibrium of CuEXC between proteins and EDTA had been achieved. After this time, the mixture (400 µL) was transferred to the Amicon® Ultra-4 centrifugal filter with a 30-kDa cut-off membrane to be centrifuged for 45 min at 2000 g at 4 °C. For CuEXC determination,

100 µL of the filtration were transferred to another centrifuge tube and diluted HNO_3 (φ = 2 %) was added up to 1 mL. The rest of the filtrate was kept for isotopic analysis.

To save sample (as a limited amount was available), total Cu determination was carried out after sample preparation (Cu isolation) for isotopic analysis, as mentioned below. Total Cu determination in the reference materials was carried out after 30 times dilution of the original material with HNO_3 with a volume fraction (φ) of 2 %.

For Cu isolation before isotopic analysis, both the total Cu fraction (bulk serum) and the CuEXC fraction (after ultrafiltration step) were treated in the same way. The first step for Cu isolation was sample digestion. For that purpose, 0.15 - 0.5 mL of bulk serum and the remaining amount of the CuEXC fraction (around 0.15 mL, exact sample volumes depended on the total sample amount available), were mixed in quartz vials with 3 mL of 14 mol L^{-1} HNO_3 and were digested in a microwave system following the program recommended by the manufacturer for blood. The digest was then transferred to Teflon Savillex® vials to be evaporated until almost dryness at 95 °C using an EVAPOCLEAN® system.

Once the digestion and the evaporation were concluded, Cu isolation was carried out using a Triskem Cu specific resin following the method proposed by Miller et al.,[37–39] which was further validated for serum samples in our research group.[12] The evaporated fraction was redissolved in 500 µL of 12 mol L^{-1} HCl and heated at 120 °C in a closed Teflon Savillex® vial for 24 h, to ensure that Cu is found in its chloride form at the highest oxidation state. Then, the sample was evaporated until almost dryness in the EVAPOCLEAN® system and redissolved in 4 mL of 5 mmol L^{-1} HCl. This process was repeated twice to ensure that excess HCl was eliminated. The chromatographic columns were prepared by adding 0.5 mL of Triskem Cu specific resin into the Bio-Rad Poly-Prep columns. A piece of cotton was used as stopper on the top of the resin. The

resin was soaked in EtOH aqueous solution (φ = 20 %) overnight before its use as recommended by the manufacturer. The chromatographic separation was carried out as described elsewhere for serum samples[12] and the protocol is summarized in **Table S1** (Electronic Supplementary Information). The Cu fraction obtained was then evaporated until almost dryness at 95 °C in an EVAPOCLEAN® system and redissolved in 50 µL of diluted HNO_3 (φ = 2 %).

For total Cu determination using the ICP-MS, the fraction of total Cu was spiked with Ni at a final concentration of 50 µg L^{-1}, as this element was chosen as internal standard. For isotopic analysis, both fractions (total Cu and CuEXC) for each sample were divided into 6 groups, depending on the total Cu concentration. The Cu concentration was adjusted with diluted HNO_3 (φ = 2 %) to 50, 100, 150, 200, 350, and 1000 µg L^{-1}, matching with the standard used for the bracketing correction method to avoid additional mass bias. Moreover, the samples were spiked with Ni (NIST SRM 986) at a final concentration of 2 mg L^{-1}, to achieve enough signal, as this element was selected as internal standard for mass bias correction. A visual schematic of the procedures for sample preparation are shown in **Figure S1** (ESI).

2.4. Measurement protocol

2.4.1. Determination of both exchangeable Cu and total Cu.

The experiments related to evaluation and validation of CuEXC protocol were conducted both at the University of Zaragoza (using the Perkin Elmer Nexion 300X) and at the IPREM in Pau (using the Perkin Elmer DRC II) using RMs L-1 and L-2. The ICP-MS devices were optimized daily for maximum stability and sensitivity. The measurement for both total Cu and CuEXC were carried out in continuous mode under the conditions shown in **Table 1.** Quantitative results were obtained using a calibration curve (one blank solution, five calibration points, using single-element Cu standard for

the analysis) between 5 and 100 μg L^{-1}. In all cases, three replicates per sample were measured. The results obtained by both laboratories were further compared in order to verify the quantification method (see **Table 3**).

CuEXC determination in real samples was carried out in continuous mode with the ELAN DRC II ICP-MS using the reaction cell with NH_3 gas, to minimize spectral overlap. The instrument and data acquisition parameters used during these analyses are gathered in **Table 1**. For quantification, a calibration curve between 1 and 10 μg L^{-1} Cu (consisting of one blank solution and four calibration points, prepared using a single-element Cu standard) was used.

As the sample volume was limited and in order to keep the maximum of this volume for MC-ICP-MS analysis, total Cu determination in real samples was performed using time-resolved analysis (TRA) *via* the direct μ-injection method developed and evaluated by our research group for total Cu determination.[12] In this case, the sample introduction system consisted of a PFA-ST MicroFlow Nebulizer with a capillary sampling tube of 0.25 mm internal diameter mounted in a peristaltic pump. This capillary was connected to a PVC tubing (orange-blue tube, ID 0.25 mm x OD 2.07 mm from Glass Expansion), which in turn was connected to another additional capillary tube of 0.25 mm ID in the other end (external capillary) functioning as a sample probe. During the analysis, except for the time of sample introduction, a diluted HNO_3 solution (φ = 2 %) was continuously aspirated to rinse. For that purpose, the external capillary was introduced into the vial with the HNO_3 solution. For sample injection, the external capillary was disconnected from the peristaltic pump tube for a few seconds and, after that, 1 μL of sample introduced into the air space formed in the peristaltic pump tube with the help of an electronic micropipette. After the injection, it was necessary to wait for a few seconds before reconnecting the external capillary to the peristaltic pump

tube to obtain a well-defined peak (**Figure 1a**). For quantification, a Cu calibration curve between 0.1 and 20 mg L^{-1} (one blank solution, seven calibration points, using single-element Cu standard for the analysis), spiked with Ni as IS (at a final concentration of 50 µg L^{-1}), was measured and the quantification was verified analysing the serum RMs L-1 and L-2 (*cf.* section 2.3). The instrument and data acquisition parameters used for total Cu determination are gathered in **Table 1.** Peak area was used for quantification, and three replicates per sample were carried out.

2.4.2. Determination of Cu isotope ratios

The optimization of the MC-ICP-MS instrument coupled to a Teledyne CETAC Aridus3 Desolvating Nebulizer System was carried out in continuous mode with a solution of 50 µg L^{-1} of both Cu and Ni, aiming at maximum sensitivity and stability. The low-resolution mode was used to maximize the sensitivity, taking into account that Cu had been isolated from the samples before the analysis, thus minimizing any risk of spectral overlap. The instrumental parameters selected are summarized in **Table 2**. For isotopic analysis, sample introduction was carried out in a similar way as for the determination of total Cu (*cf.* section 2.4.1), but, in this case, 3 µL of pre-treated sample were injected instead of 1 µL, and the sample was introduced by self-aspiration (no peristaltic pump was used). For that purpose, the sample was directly injected into the nebulizer tubing (1.3 mm OD x 0.25 mm ID) with an electronic micropipette. Under these conditions, the signal duration was around 20 - 30 seconds (See **Figure 1b**), depending on the Cu concentration. Between injections, a diluted HNO$_3$ solution (φ = 2 %) was aspirated for cleaning. Five measurements were carried out per sample following a standard-sample-standard bracketing sequence using NIST SRM 3114 as standard.

Before the measurements, the MC-ICP-MS coupled to the desolvating nebulizer system was optimized, and the performance was checked in continuous mode. For this purpose, a mixture of NIST SRM 3114 (Cu) and NIST SRM 986 (Ni) at a final concentration of 2 mg L^{-1} for each element was used. The $^{65}\delta$ value was calculated using the self-bracketing method, *i.e.*, the standard is considered as standard and as sample at the same time, except for the first and the last measured, which are only considered as standards. Using this strategy, the $^{65}\delta$ value expected is 0. Moreover, Ni was used for mass bias correction. The average $^{65}\delta$ value obtained was 0.00 ± 0.18 ‰ (N=10, 2SD).

All Cu isotopic ratio calculations were carried out with the Linear Regression Slope (LRS) method. With this method, the values measured for both Cu isotopes are confronted to obtain a linear regression. The Cu isotope ratio is calculated as the slope of this regression.[40] For mass bias correction, a combination of internal and external normalization was used. Internal normalization was carried out according to the exponential model described by Maréchal *et al*.[41] using Ni. The factor to correct the Cu ratio was calculated by determining experimentally the $^{62}Ni/^{60}Ni$ ratio *via* LRS and using the certified $^{62}Ni/^{60}Ni$ ratio ($^{62}Ni/^{60}Ni$ = 0.138600 ± 0.000045). External normalization was carried out using the standard-sample-standard bracketing sequence with the NIST SRM 3114. More information about the equations used is presented as ESI.

Five replicate ratios were obtained, and the mean value was used as representative result. The uncertainty was calculated as two times the standard deviation.

2.5. Machine learning protocol

To develop the ML models, we have used the Spyder open source scientific environment for python[42] and the Keras interface for the TensorFlow open-source

machine-learning platform.[43] We have selected Artificial Neural Networks (ANNs)[44] as the basis for our ML solutions, because, for small problems, it is one of the most powerful ML approaches and it is included in all the ML environments. Moreover, we will use an ensemble of ANNs to reduce the variance of our model and to enrich its output with an additional value that measures the uncertainty of each prediction. As explained by Bhatt *et al*.,[45] measuring the uncertainty of a prediction is a form of transparency that allows for better understanding of predictions and for improved decision-making. However, very few works to date measure and analyze the uncertainty of the predictions of the ML models developed. We will carry out such analysis and demonstrate its utility to select the input parameters for our models, and to achieve a better understanding of the results finally obtained.

ML is based on the "learning from data" paradigm. In our case, the main limitation is that the dataset is very small. As explained before, we have information of only 56 volunteers from the Hospital Miguel Servet (Zaragoza, Spain). When the data set is small, the greatest risk is falling into overfitting. Our goal is to create a system that generalizes well to unseen examples. However, if we use a large network, and we overtrain it, our model may overfit the training data, memorizing all the different inputs, and generating the perfect output for them. However, this is not a good solution, since overfitting models are not able to generalize their results for other inputs unseen during the training process.

For model selection, we carried out a grid search exploring four parameters: the number of hidden layers, the number of neurons per layer, the learning rate, and the maximum number of epochs to train.

During the model explorations, we found several hyperparameters that achieve similar, very high accuracy results. This indicates that the data collected contains sufficiently

16

clear and relevant information for the models to use. For our analysis, we could use any of the models that provide equivalent results. We selected the smaller one, which is always a good criterion to prevent overfitting. Regarding the number of epochs, we stopped training as soon as the model reached stable accuracy results for the training set, even though the loss function that guides the training process indicated that the fit could be improved.

The final selected model is a small ANN classifier with three levels of four neurons each that use Rectified Linear Unit (RELU) as activation function, and a final output layer with two neurons that use Softmax activation function to generate a probability distribution for the two outputs: WD / Non-WD. These outputs can be interpreted as the probabilities of suffering or not suffering from WD. The model was trained 2000 epoch with a 0.001 learning rate. As inputs, we started using only the values of Cu total concentration and CuEXC concentration for the samples, and after that we explored the effect of including additional inputs. We normalized each value using the standard score to improve the training process. To reduce the variance of the output, we have trained 50 different models with the same setup. Since the training process started from different random initial points, each one of these models is different from the others. Our final model is an ensemble of these 50 NNs, and its classification output is the average of the 50 individual outputs. The internal variations among the 50 models, together with the probabilities generated by the softmax layer (the last layer of the ANN, which generates its final outputs) of the ANN are used to compute the uncertainty values.

The code used to define the model and train it, and the code used to generate the uncertainty metric is published in a GitHub repository.[46]

3. Results and discussion

For the elemental determination of Cu, *i.e.*, non-isotopic analysis, all the blanks (including with EDTA and KCl) were below 1 µg L^{-1} Cu. For the isotopic analysis, the δ^{65}Cu was 0.23 ± 0.21 ‰ for the solutions using diluted HNO$_3$ (φ = 2 %), and 0.42 ± 0.22 ‰ for the CuEXC solution (EDTA+KCl). They were all subtracted when calculating both Cu concentrations and isotope ratios.

3.1. Evaluation of the method for CuEXC determination

To determine CuEXC, careful sample preparation is mandatory. To evaluate the method and validate the results before real sample analyses were conducted, an interlaboratory study was performed. The same serum reference material (L-1 and L-2) were subjected to the CuEXC separation procedure (as explained in section 2.3.), and subsequent analyses in two different laboratories, one in Pau (France) and the other in Zaragoza (Spain), were carried out. In addition to CuEXC, total Cu was also determined, and REC was calculated following equation (1). The results obtained are shown in **Table 3.** The total Cu concentration values observed in both laboratories differ (higher levels were obtained in Zaragoza, particularly for RM L-1), but both concentrations are within the concentration range provided by the reference materials. These differences may be due to small variations in the sample preparation because the reference material is freeze-dried and it is necessary to reconstitute it (in principle, Pau lab facilities are less prone to suffer from contamination issues). CuEXC values follow the same trend (a bit higher values are obtained in Zaragoza), but the REC values are rather similar in both cases, which points to minor differences in the reconstitution step, as commented before. The results were overall considered as fit-for-purpose, and analysis of the real samples was undertaken next.

Once the CuEXC separation protocol was evaluated in terms of Cu recovery, mass fractionation during the whole analytical procedure (including both CuEXC separation and Cu isotopic isolation) was investigated. To verify that no mass fractionation occurs during these sample preparation steps, three aliquots of NIST SRM 3114 were subjected to both procedures consecutively, as detailed in section 2.3. After the two consecutive protocols, each aliquot was adjusted to a different Cu concentration for analysis: 250, 500, and 1000 µg L^{-1}.

Next, isotopic analyses in TRA mode (*cf.* section 2.4. and **Figure 1b**) were performed, using a direct µ-injection method. 10 replicates per sample were measured, and the results were expressed as $^{65}\delta$ value, with the standard-sample bracketing method, where the standard is the NIST SRM 3114, and the sample is the NIST SRM 3114 after suffering the two consecutive sample preparation procedures (CuEXC separation and Cu isolation). If no isotope fractionation is present, the recorded $^{65}\delta$ value should be 0. The results obtained are shown in **Table 4** and external precision is expressed as 2SD in ‰. As can be seen from this table $^{65}\delta$ values were around 0, meaning that no significant fractionation occurs during both procedures.

These results are in accordance with previous investigations. In fact, a similar method was also validated by Lauwens *et al.*,[32] but in that case, a different procedure was carried out for Cu isolation using the AG-MP-1 resin. The differences between the performance of these two resins (AG-MP-1 and Trisken) are discussed in ref. 12. and will not be further contrasted herein.

3.2. Analysis of the samples for total copper, exchangeable copper, and REC

To interpret the results, the samples were divided into four different groups: controls (healthy patients, C), patients with hepatic disorders different from WD (HD), newborns/infants (NB) and WD patients (WD).

3.2.1. Total Copper

Total Cu determination was carried out after the procedure for Cu isolation with an ion-exchange resin (see section 2.3.), to save sample for other tests because the amount of sample available was limited. Analyses were carried out as explained in section 2.4.1., and the results obtained are displayed in **Figure 2a**.

The Cu concentration obtained for WD patients, regardless of the treatment, ranged between 49 and 498 µg L^{-1} (mean value = 147 µg L^{-1}). For newborns and infants, the concentration found varied between 134 and 626 µg L^{-1} (mean value = 373 µg L^{-1}). For controls and patients with other liver disorders, values between 414 and 2901 µg L^{-1} (mean value = 1128 µg L^{-1}), and between 625 and 2022 µg L^{-1} (mean value = 1410 µg L^{-1}), respectively, were observed. According to these results, WD patients show lower Cu levels than the rest of the groups tested in this work. However, some overlapping between WD patient Cu levels and those of the rest can be observed, particularly with the levels found for newborns and infants, as expected.[12,15] This confirms that total Cu determination is useful, but it is not a selective biomarker for WD diagnosis.[13]

3.2.2. Exchangeable Copper

The analyses to determine CuEXC were carried out as described in section 2.4.1. The results are plotted in **Figure 2b**. No differences were found among the four groups regarding CuEXC. The CuEXC concentration found in WD patients ranged between 18.9 and 88.6 µg L^{-1} (mean = 37.1 µg L^{-1}), overlapping with the CuEXC levels found in

the controls, patients with other liver disorders and newborns/infants, which showed

ranges of 32.1 - 82.6 µg L^{-1} (mean = 54.8 µg L^{-1}), 42.7 – 124.0 µg L^{-1} (mean = 59.1 µg

L^{-1}) and 13.9 – 53.6 µg L^{-1} (mean = 36.7 µg L^{-1}), respectively. The results obtained in

this study for the controls agree with the reference values established by El Balkhi *et

al.*[25] for CuEXC for WD patients, although the number of samples in both studies is

small.

It is necessary to remark that all WD patients analyzed in this study were under

treatment, except for one patient. Moreover, a patient who was suspected of not

following the treatment correctly present the highest CuEXC level (88.6 µg L^{-1}), which

again agrees well with the report by Guillaud *et al.*[19]

On the other hand, the WD patient without any prior treatment shows a CuEXC level

of 38.4 µg L^{-1}, which seems to be low for a patient just diagnosed. However, this patient

was asymptomatic at the moment of blood collection, which agrees with the results

reported by Guillaud *et al.*[19] In view of the results, it is possible to state that CuEXC

determination by itself is not enough for reliable WD diagnosis. However, it could be

used to follow up the evolution of the disease. In any case, further studies with a larger

population are needed to verify these aspects. As stated before, this is a typical

problem when dealing with rare diseases, as it is never easy to collect a large number

of samples.

3.2.3. Relative exchangeable copper

The results obtained show differences between REC values for WD patients and for

the other groups of persons subjected to study. WD patients show higher REC than

the other groups tested (see **Figure 2c).** In particular, the REC obtained for WD

patients fall in the range between 17.8 and 64.7 %, while the rest of the groups show

REC values between 2.3 and 14.9 %. Therefore, with the population under study, a cutoff value in the range of 15 to 17 % REC would enable detection of WD with both selectivity and specificity values of 100%. The current work indicates that REC seems to be a useful parameter with the capacity to identify WD patients from other groups. It is even possible to differentiate WD patients from healthy newborns, which indicates that this biomarker could be used in neonatal screening programs.

Similar conclusions have been drawn in the literature, although the cutoff value is not clearly defined yet. In their first study with REC, El Bakhi et al.[18] found a cutoff of 18.5 % providing a 100 % sensitivity and specificity for distinguishing WD patients from controls, wild-type homozygous and heterozygous. Guillaud et al.,[19] on the other hand, could also distinguish WD patients from controls (patients with hepatic disease different to WD) using the REC. A cutoff of 18.5 % was set taking into account only WD patients sampled at the moment of diagnosis or who failed to respond to treatment because of non-compliance. However, if WD patients on stable condition under medical treatment were also included, the cutoff decreased to 14 % and provided one false positive. Moreover, they found that the REC was not influenced by the presence of hepatic disorders, as such patients were indistinguishable from healthy controls. Also, no differences in the REC values were observed between WD patients with/without cirrhosis. Trocello et al.[27] used the REC for family screening and showed that, by setting a cutoff of 15 %, it was possible to differentiate WD patients (carrying 2 mutations in the ATP7B gene) from the subjects without WD (ATP7B heterozygous carriers, and subjects with no identified mutation in the ATP7B gene) with 100 % of sensitivity and specificity. In a recent study, subjects with alcoholic cirrhosis showed REC values ≤ 19 %, except for two patients showing REC values of 21 and 25 %.[32]

However, no genetic tests were carried out in that work to investigate the potential disease of such patients.

On the other hand, in studies with animals, where it is possible to control much better almost all the external parameters, Schmitt *et al.*[17] found that it was possible to discriminate Long-Evans Cinnamon rats (LEC, with an ATP7B mutation causing WD) from Long-Evans rats (LE, without any mutation) with a sensitivity of 97.3 % and specificity of 100 % setting a cutoff of 19 %. They also showed that the REC is not influenced by the presence of liver damage or by the Cu intake. In another animal model, Heissat *et al.*[28] set a cutoff at 20 % for achieving a sensitivity and specificity of 100 % between WD rats and wild-type rats.

To sum up, our results agree well with the previous studies, further confirming that the REC value can be a sensitive and specific tool for diagnosing WD. However, further studies (with a higher number of both WD and controls) are needed to be able to set a reliable and constant cutoff for the REC (which seems to be in the range between 14 and 20, according to all the literature) and to know the limitations of this biomarker.

3.3. Cu Isotopic analysis

In order to further understand the evolution of the disease and to complement the previous biomarkers, Cu isotopic analyses were carried out.

To better interpret these analyses, the samples were divided into six different groups in this case: controls (healthy patients, C), patients with hepatic disorders different from WD (HD), healthy infants and newborns (NB), WD under treatment with a chelator (WD-C), WD under treatment with a Zn salt (WD-Zn) and one sample from a WD patient obtained at the moment of diagnosis, without any treatment (WD).

Next, these samples were analyzed for Cu isotopic composition in both bulk serum and the CuEXC fraction. Five replicates per sample were carried out, following the bracketing sequence (*cf.* section 2.4.2). **Figure 3** shows the results of $\delta^{65}Cu$ (‰) obtained in both bulk serum (dark colors) and CuEXC fraction (light colors) for the samples under investigation.

The isotopic composition in bulk serum for control subjects show $\delta^{65}Cu$ values around 0 with an average $\delta^{65}Cu$ value of 0.08 ± 0.27 ‰ (expressed as $\bar{x}$ ± SD). Patients with other hepatic disorders show an average $\delta^{65}Cu$ of -0.62 ± 0.51 ‰. Results obtained for newborns show $\delta^{65}Cu$ values around 0 or positive, with an average value of 0.21 ± 0.33 ‰. For WD patients under treatment with a chelator, a negative $\delta^{65}Cu$ value was found, with an average value of -0.58 ± 0.56 ‰, while WD patients under treatment with Zn salts show positive delta values, with an average value of 0.53 ± 0.47 ‰.

It is interesting to mention that in this latter group there is a patient showing a relatively large negative $\delta^{65}Cu$ (-0.36 ‰). This sample corresponds to the same patient that showed a $\delta^{65}Cu$ of -1.88 ‰, in the group of WD under chelator treatment (value outside the boxplot for WD-C) in a previous sampling when the patient was under treatment with both a chelator and a Zn salt. That means that when the patient stopped the treatment with the chelator keeping only the Zn salt treatment, the $\delta^{65}Cu$ shifted to heavier values. However, it was still negative. The sample of the patient at the moment of the WD diagnosis (she/he was asymptomatic) shows a $\delta^{65}Cu$ of 0.68 ‰.

The results of $\delta^{65}Cu$ obtained in bulk serum for controls were consistent with those previously reported in the literature.[32,47,48] The results obtained for patients with other hepatic disorders seem to be lighter than the results obtained for the control patients and they agree with previously published results.[32,48,49] For newborns, the results

obtained are similar to the values obtained for the controls, and also they agree with

previously reported data.[12] In the same way, the results obtained for WD, regardless

of treatment, are in agreement with the results reported before.[12] The δ^{65}Cu values are

lighter for WD patients treated with a chelator compared to WD patients treated with

Zn salts, and they are comparable with the δ^{65}Cu obtained for patients with other

hepatic problems. That is coherent with the hypothesis of ^{63}Cu being preferably

accumulated in the liver, and the negative delta values being due to Cu released from

the liver, either by the treatment with a chelator or by liver disorders, as postulated in

a previous publication.[12]

On the other hand, the isotopic composition in the CuEXC fraction shows average

δ^{65}Cu values of 0.38 ± 0.28 ‰ for control subjects. For patients with hepatic disorders

(non-WD) the average value is -0.70 ± 0.58 ‰, while the newborns show a mean value

of -0.11 ± 0.99 ‰. For WD patients under chelating treatment, the mean value of δ^{65}Cu

is -0.57 ± 0.43 ‰, while for the patient under Zn salt treatments, this value is 0.58 ±

0.53 ‰. The WD patient without any treatment (at the moment of WD diagnosis) shows

a positive delta value of 0.72 ‰.

Results of δ^{65}Cu obtained in the CuEXC fractions seem to follow the same trend as

δ^{65}Cu obtained in bulk serum. That means lighter values were observed for patients

with hepatic problems and WD patients under chelating treatment, while heavier values

were found for controls, WD patients under Zn salt treatment, and for the only sample

of WD without treatment and asymptomatic. In the case of newborns/infants, there is

a very high variability in comparison with bulk serum.

To enable direct comparison of the results obtained in both fractions (bulk serum and

CuEXC) for the same sample, the Δ^{65}Cu values were calculated following the equation

(2) and are plotted in **Figure 4.**

$$\Delta^{65}Cu = \delta^{65}Cu_{EXC} - \delta^{65}Cu_{bulk\ serum} \tag{2}$$

In this figure, the only group that follows a clear trend is the control group. In this group, it is possible to observe that, in all cases, the isotopic composition of CuEXC fraction shifts to heavier values compared to the isotopic composition in bulk serum. Meanwhile, for the rest of the samples under investigation, a high variability between them prevents obtaining a clear conclusion.

The average $\Delta^{65}Cu$ value for control subjects was 0.36 ± 0.18 ‰. For patients with other hepatic disorders, the average $\Delta^{65}Cu$ was -0.08 ± 0.36 ‰. Newborns and infants, show an average value of $\Delta^{65}Cu$ -0.32 ± 0.83 ‰. WD patients, show average values of 0.01 ± 0.59 ‰ and 0.06 ± 0.68 ‰ for patients under chelator treatment or under treatment with Zn salts, respectively.

The mean value obtained for controls is in agreement with the result obtained by Lauwens et al.[32] (0.41 ± 0.15 ‰, N=7). On the other hand, Tennant et al.,[50] via computational calculation, obtained that $\delta^{65}Cu_{albumin}$ should be around 60 % heavier than $\delta^{65}Cu_{bulk\ serum}$, assuming that serum consists of 10 % albumin and 90 % ceruloplasmin. In that study, healthy patients (controls) show an average $\delta^{65}Cu_{albumin}$ 37 % heavier than in bulk serum. Our results show the same trend, but the differences between $\delta^{65}Cu_{EXC}$ and $\delta^{65}Cu_{bulk\ serum}$ are not so significant, most likely because not all CuEXC is bound to albumin. Cu may also be loosely bound to other proteins and to other low molar mass molecules, such as amino acids,[32] which were not taken into account for the computational calculation referred above. Moreover, the nature of these compounds might be highly variable among subjects belonging to the same group.[31,32]

The mean value obtained for patients with hepatic disorders (non-WD) is also in agreement with the value reported previously (-0.10 ± 0.23 ‰, N=14).[32] Despite this fact, it was not possible to find any clear trend, as both positive and negative $\Delta^{65}Cu$ were observed. The same occurs for WD independently of the treatment, and for newborns and infants where the variability in the results is high.

To try to explain the variability in the $\Delta^{65}Cu$, it is important to keep in mind that the liver is the principal source of serum proteins. Liver disorders or liver immaturity (in the case of newborns) can affect hepatic metabolism, causing changes in the levels of some proteins and in the coordination environment of Cu, thus affecting to the isotopic composition. The isotopic fractionation in general is influenced by the ligand coordination and by the oxidation state.[51,52] Generally, heavy isotopes prefer bonding with ligands with stronger electronegativity (O > N > S). Particularly, if the metal is in its highest oxidation state. Therefore, it is expected to observe a lighter isotopic composition in Cu bound to cysteine or methionine (Cu-S) than in Cu bound to histidine (Cu-N).

In control patients, it is possible to see a clear trend for $\Delta^{65}Cu$, probably because not all healthy bodies are equal, but they tend to work similarly. However, for the rest, since Cu can bond to different molecules depending on the disorder and on its state, it seems not feasible to appreciate a clear trend above the variability. Information about patients in terms of the levels of the different molecules that can bound Cu (ceruloplasmin, albumin, alpha-2-macroglobulin, etc.) was not available for this study, which could enhance the understanding about this topic.

In any case, it becomes clear that, while Cu isotopic analysis can provide relevant information related to the evolution of the disease (*e.g.*, once one knows that somebody is affected by WD, Cu isotopic analysis can provide information on how the

disease is progressing), it is hardly feasible to use it as the only diagnostic tool, as too many different factors affect the Cu ratios obtained (*e.g.*, treatment, existence of other hepatic disorders, etc.).

3.4. Machine Learning

As discussed in the Introduction section, applying ML strategies to our data set can help in developing a more efficient way to diagnose WD. Section 2.5. describes in detail how our model was built and trained. In short, we have followed a cross-validation approach to train our ANN models. Since we have 56 data available, we have divided them into 4 subsets, and we have performed four training runs using in each one 3 subsets for training and one for the test. With this approach, we have obtained test results for all input data.

Figure 5a presents the results of this approach for the 56 inputs (values of Cu total concentration and CuEXC concentration for the samples) of the data set. The prediction of the ANN is correct if the output assigns the highest probability value to the correct class. In this case, it is not correct for two of the inputs. This represents a 96.4% accuracy, computed as the number of times that the correct class is selected divided by the number of predictions (54/56), multiplied by 100. However, as explained before, accuracy is not our only concern. We also want to measure the uncertainty of the model. This can help us to understand the predictions, identify how to improve them, and to decide whether we can trust them, or else, additional analyses are needed.

For this purpose, we use the predictive entropy to measure the uncertainty of the output using equation (2), where x is the input, y is the output, c_k are the ground-truth class label, $p(y = c_k|x)$ are the probabilities assigned to each category, K is the number of

categories, and T the number of different outputs (50 in our case, since we have 50 different models working in parallel).

$$H(y \mid x) := - \sum_{k=1}^{K} (\tfrac{1}{T}\sum_{t=1}^{T} p(y = c_k|\mathsf{x}))\log(\tfrac{1}{T}\sum_{t=1}^{T} p(y = c_k|x)) \quad (2)$$

This equation measures the uncertainty inherent to the variable's possible outcomes. If for a given input, the prediction is very clear, then $H(x)$ will we very small. This will happen if all the 50 models assign almost 1 to one of the categories, and almost 0 to the other one. On the contrary, if the probabilities are not clear, or the variance in the predictions of the models is very high, the uncertainty will be high. For example, if some ANNs assign the maximum probability to one category, while others assign it to the other category, or the models assign similar probabilities to both categories, the predictive entropy will increase.

As can be seen on **Figure 5b**, the predictive entropy, uncertainty in the figure, for the two wrong predictions is very high, but it is also high for other inputs which predictions were correct. This means that we should not trust the model output for that high uncertainty data, even when the model selected the right class. In fact, for a binary classifier like this one, a random model would be correct in half of the cases. To deploy the ML model as an automated diagnostic tool, robust predictions are needed. To this end, in practice, a threshold can be set and, if the uncertainty is above it, discard that output, and ask an expert to analyze that particular case. After such data has been analyzed by experts, they can be included in the data set, and iteratively, they can be used to further improve the model.

We have analyzed the data to understand why some inputs exhibit high uncertainty values. The main reason is that some of the newborn samples show Cu total and CuEXC concentration values that are close to some of the samples of WD cases used

during training. In fact, if we compute the two-dimensions Euclidean distance among all samples, the fact is that for samples 31, 34 and 35, which belong to non-WD patients (newborns), the closest sample originates from a WD patient, and for sample 40, which is a WD sample, the closest one belongs to a newborn.

In one of the cases, sample 26, the reason is different. This patient was affected by a hepatic disorder and showed a very high CuEXC concentration, which is out of the range in comparison with the others (highest value in **Figure 2b**, above 120 µg L^{-1}). When such data are in the test set, the model has not seen anything like it during training, and for that reason, it generates a very high uncertainty. This is the desired behavior for an automated diagnostic tool. In a medical application, the model should alert for expert review when it encounters data that are very different from those used during training.

Although we have obtained a good accuracy, 96.5% (see **Figure 5**), we wanted to further improve our models to reduce their uncertainty. To do so, we added more information to the inputs. **Figure 6** presents the probabilities and the entropy of our model when we train it adding the REC value to the inputs.

As can be seen in the figure, the results have clearly improved. In this case the model properly classifies all the cases (100% accuracy), and the probabilities assigned to the correct class are higher than before, being very close to the value 1 in most cases. As for the uncertainty (see **Figure 6b**), the average value has been reduced from 0.175 to 0.077. Hence, including the REC value helps to develop a more robust model. This is interesting, since the REC values are calculated from the other two inputs (see equation (1)), but the model improves when this parameter is also added. The only case with no improvement is number 26, which, as explained before, corresponds to a

data out of the range of the others for the CuEXC value. This high uncertainty value is useful to identify this anomaly.

Continuing with this analysis, we added two more input values: $\delta^{65}Cu_{bulk\ serum}$ and $\delta^{65}Cu_{Exch}$. **Figure 7** presents the results. In this case, the additional data increases the uncertainty of the model. Hence, it is not recommendable to use it, because the model output is more robust without it. This can be already anticipated since the $\delta^{65}Cu$ values strongly depends on the treatment that patient is following (see **Figure 3**) and, thus, cannot help in differentiating WD patients from the rest.

The uncertainty metric clearly helps to choose which information should be used as model inputs, but it can also be useful to detect problems with the predictions. As explained before, we have normalized all the inputs before training. To use the model properly, it is important to normalize the inputs using the same normalizing factors always. However, this may be forgotten by mistake. We have analyzed what would happen in that case by entering the test inputs without such normalization. The result is that our classifier fails 30% of the tests. This is normal, if the inputs are not in the correct format, the model cannot do their computations properly. The interesting result is that the uncertainty metric increased by a factor of 12, thus clearly highlighting that these predictions were very unreliable. Therefore, thanks to this metric, it is possible to detect anomalous situations, and try to understand and solve them, thus highly improving the confidence in our results.

Conclusion

This work presents a comprehensive approach of what can be achieved nowadays *via* ICP-MS analysis of serum samples for the study of WD, combining both isotopic and

species-specific Cu information, The results demonstrate that this technique is able to provide key information for diagnostic and follow up of WD.

While Cu isotopic analysis *via* MC-ICP-MS can provide data for better understanding of the disease and for follow-up of WD patients, its contribution to diagnosis seems to be of limited value: Cu delta values for WD patients depend on the treatment followed and overlap with those observed for patients with other hepatic issues, regardless of the Cu fraction subjected to analysis.

On the other hand, it seems that a simple fractionation approach to determine total Cu and CuEXC fraction, together with the calculation of the REC and the implementation of a ML model can help in providing a more efficient WD diagnosis, which could eventually help in establishing newborn screening programs. Moreover, this information can be obtained in an easy way, enabling hospitals to perform these determinations routinely using a "conventional" (meaning quadrupole-based) ICP-MS, or even other atomic techniques that are sufficiently sensitive (*e.g.*, graphite furnace atomic absorption spectrometry), even though it is hard to compete with the high sample throughput of ICP-MS in this context.

Regarding the application of ML, the analysis of the uncertainty is one of the main contributions of this work. In the literature, you can find one or several ML models for every existing dataset. However, the focus is always to improve accuracy, but, for medical diagnosis, accuracy is not enough. Uncertainty estimation is needed to build a robust automated diagnostic tool. This work shows how the use of an uncertainty metric can help to better understand the results obtained during training, identifying anomalies or ambiguous data. In this way, if the model generates an output with a very small uncertainty, it can be interpreted that the model was trained with similar data and the output is reliable. On the other hand, if the uncertainty is high, it can be anticipated

that the prediction is not fully reliable, and an expert should check the data and make the final decision.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1   M. Schaefer and J. D. Gitlin, *Am. J. Physiol.-Gastrointest. Liver Physiol.*, 1999, **276**, G311–G314.

2   S. Lutsenko and M. J. Petris, *J. Membr. Biol.*, 2002, **191**, 1–12.

3   J. D. Gitlin, *Gastroenterology*, 2003, **125**, 1868–1877.

4   A. Ala, A. P. Walker, K. Ashkan, J. S. Dooley and M. L. Schilsky, *Lancet*, 2007, **369**, 397–408.

5   I. Mohr and K. H. Weiss, *Ann. Transl. Med.*, 2019, **7**, S69.

6   E. Roberts and M. L. Schilsky, *Hepatology*, 2003, **37**, 1475–1492.

7   M. Bost, G. Piguet-Lacroix, F. Parant and C. M. R. Wilson, *J. Trace Elem. Med. Biol.*, 2012, **26**, 97–101.

8   S. M. Kenny and D. W. Cox, *Hum. Mutat.*, 2007, **12**, 1171–1177.

9   I. J. Chang and S. H. Hahn, in *Handbook of Clinical Neurology*, ed. P. Vinken and G. Bruyn, Elsevier, Amsterdam, 2017, vol. 142, ch. 3, pp. 19–34.

10   E. A. Roberts and M. L. Schilsky, *Hepatology*, 2008, **47**, 2089–2111.

11   C. D. Romero, P. H. Sanchez, F. L. Blanco, E. R. Rodríguez and L. S. Majem, *J. Trace Elem. Med. Biol.*, 2002, **16**, 75–81.

12   M. C. García-Poyo, S. Bérail, A. L. Ronzani, L. Rello, E. García-González, B. Lelièvre, P. Cales, F. V. Nakadi, M. Aramendía, M. Resano and C. Pécheyran, *J. Anal. At. Spectrom.*, 2021, **36**, 968–980.

13   F. Woimant, N. Djebrani-Oussedik and A. Poujois, *Ann. Transl. Med.*, 2019, **7**, S70.

14   T. Müller, S. Koppikar, R. M. Taylor, F. Carragher, B. Schlenck, P. Heinz-Erian, F. Kronenberg, P. Ferenci, S. Tanner, U. Siebert, R. Staudinger, G. Mieli-Vergani and A. Dhawan, *J. Hepatol.*, 2007, **47**, 270-276.

15   M. Aramendía, L. Rello, M. Resano and F. Vanhaecke, *J. Anal. At. Spectrom.*, 2013, **28**, 675–681.

16   M. Resano, M. Aramendía, L. Rello, M. L. Calvo, S. Bérail and C. Pécheyran, *J. Anal. At. Spectrom.*, 2013, **28**, 98–106.

17   F. Schmitt, G. Podevin, J. Poupon, J. Roux, P. Legras, J.-M. Trocello, F. Woimant, O. Laprévote, T. H. NGuyen and S. E. Balkhi, *PLoS ONE*, 2013, **8**, e82323.

18   S. El Balkhi, J.-M. Trocello, J. Poupon, P. Chappuis, F. Massicot, N. Girardot-Tinant and F. Woimant, *Clin. Chim. Acta*, 2011, **412**, 2254–2260.

19   O. Guillaud, A.-S. Brunet, I. Mallet, J. Dumortier, M. Pelosse, S. Heissat, C. Rivet, A. Lachaux and M. Bost, *Liver Int.*, 2018, **38**, 350–357.

20   M. Costas-Rodríguez, J. Delanghe and F. Vanhaecke, *TrAC Trends Anal. Chem.*, 2016, **76**, 182–193.

21   F. Albarède, *Elements*, 2015, **11**, 265-269.

22   B. Mahan, R. S. Chung, D. L. Pountney, F. Moynier and S. Turner, *Cell. Mol. Life Sci.*, 2020, **77**, 3293-3309.

23   F. Vanhaecke and M. Costas-Rodríguez, *View*, 2021, **2**, 20200094.

24   F. Vanhaecke and P. Degryse, *Isotopic Analysis: Fundamentals and Applications Using ICP-MS*, Wiley-VCH, Weinheim, 2012.

25   S. El Balkhi, J. Poupon, J.-M. Trocello, A. Leyendecker, F. Massicot, M. Galliot-Guilley and F. Woimant, *Anal. Bioanal. Chem.*, 2009, **394**, 1477–1484.

26   J. M. Walshe, in *Advances in Clinical Chemistry*, ed. G. S. Makowski, Elsevier, Amsterdam, 2010, vol. 50, ch. 8, pp. 151–163.

27   J.-M. Trocello, S. El Balkhi, F. Woimant, N. Girardot-Tinant, P. Chappuis, C. LLoyd and J. Poupon, *Mov. Disord.*, 2014, **29**, 558–562.

28   S. Heissat, A. Harel, K. Um, A.-S. Brunet, V. Hervieu, O. Guillaud, J. Dumortier, A. Lachaux, E. Mintz and M. Bost, *J. Trace Elem. Med. Biol.*, 2018, **50**, 652–657.

29   M. E. del Castillo Busto, S. Cuello-Nunez, C. Ward-Deitrich, T. Morley and H. Goenaga-Infante, *Anal. Bioanal. Chem.*, 2022, **414**, 561–573.

30   N. Solovyev, A. Ala, M. Schilsky, C. Mills, K. Willis and C. F. Harrington, *Anal. Chim. Acta*, 2020, **1098**, 27–36.

31   C. D. Quarles, M. Macke, B. Michalke, H. Zischka, U. Karst, P. Sullivan and M. P. Field, *Metallomics*, 2020, **12**, 1348–1355.

32   S. Lauwens, M. Costas-Rodríguez, J. Delanghe, H. Van Vlierberghe and F. Vanhaecke, *Talanta*, 2018, **189**, 332–338.

33   W. Wang, X. Liu, C. Zhang, F. Sheng, S. Song, P. Li, S. Dai, B. Wang, D. Lu, L. Zhang, X. Yang, Z. Zhang, S. Liu, A. Zhang, Q. Liu and G. Jiang, *Chem. Sci.*, 2022, **13**, 1648–1656.

34   G. L. Donati, in *Comprehensive Analytical Chemistry*, ed. M. A. Z. Arruda and J. R. de Jesus, Elsevier, Amsterdam, 2022, vol. 97, ch. 2, pp. 53–88.

35   M. C. García-Poyo, A. L. Ronzani, J. Frayret, S. Bérail, L. Rello, E. García-González, B. Lelièvre, F. V. Nakadi, M. Aramendía, M. Resano and C. Pécheyran, *Spectrochim. Acta Part B At. Spectrosc.*, 2021, **185**, 106306.

36   Q. Hou, L. Zhou, S. Gao, T. Zhang, L. Feng and L. Yang, *J. Anal. At. Spectrom.*, 2016, **31**, 280–287.

37   K. A. Miller, PhD Thesis, University of Calgary, 2018.

38   K. A. Miller, C. M. Keenan, G. R. Martin, F. R. Jirik, K. A. Sharkey and M. E. Wieser, *J. Anal. At. Spectrom.*, 2016, **31**, 2015–2022.

39   C. Dirks, B. Scholten, S. Happel, A. Zulauf, A. Bombard and H. Jungclas, *J. Radioanal. Nucl. Chem.*, 2010, **286**, 671–674.

40   J. Fietzke, V. Liebetrau, D. Günther, K. Gürs, K. Hametner, K. Zumholz, T. H. Hansteen and A. Eisenhauer, *J. Anal. At. Spectrom.*, 2008, **23**, 955–961.

41   C. N. Maréchal, P. Télouk and F. Albarède, *Chem. Geol.*, 1999, **156**, 251–273.

42   Spyder, https://www.spyder-ide.org/ (Last accessed July 2022).

43   Keras, https://keras.io/ (Last accessed July 2022).

44   S. S. Haykin, *Neural networks and learning machines*, Prentice Hall, New York, 3rd edn, 2009.

45   U. Bhatt, J. Antorán, Y. Zhang, Q. V. Liao, P. Sattigeri, R. Fogliato, G. Melançon, R. Krishnan, J. Stanley, O. Tickoo, L. Nachman, R. Chunara, M. Srikumar, A. Weller and A. Xiang, in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, Virtual Event USA, 2021, 401–413.

46   https://github.com/JavierResano/ML-for-the-diagnosis-of-Wilson-s-Disease-via-ICP-MS

47   F. Albarède, P. Telouk, A. Lamboux, K. Jaouen and V. Balter, *Metallomics*, 2011, **3**, 926–933.

48   M. Costas-Rodríguez, Y. Anoshkina, S. Lauwens, H. Van Vlierberghe, J. Delanghe and F. Vanhaecke, *Metallomics*, 2015, **7**, 491–498.

49   S. Lauwens, M. Costas-Rodríguez, H. Van Vlierberghe and F. Vanhaecke, *Sci. Rep.*, 2016, **6**, 30683.

50   A. Tennant, A. Rauk and M. E. Wieser, *Metallomics*, 2017, **9**, 1809–1819.

51   T. Fujii, F. Moynier, M. Abe, K. Nemoto and F. Albarède, *Geochim. Cosmochim. Acta*, 2013, **110**, 29–44.

52   E. A. Schauble, *Rev. Mineral. Geochem.*, 2004, **55**, 65–111.

Table 1. ICP-MS spectrometer settings and data acquisition parameters for the determination of total Cu and CuEXC in serum samples.

	NexION 300X ICP-MS	ELAN DRC II quadrupole ICP-MS	
	Continuous mode	Continuous mode	Time resolved analysis (TRA)
Sample uptake rate ($\mu L\ min^{-1}$)	300	100	100
Nebulizer gas, Ar ($L\ min^{-1}$)	1.00	0.90	0.92
Plasma gas, Ar ($L\ min^{-1}$)	15.00	17.00	17.00
Auxiliary gas, Ar ($L\ min^{-1}$)	1.20	1.00	1.00
RF power (W)	1600	1000	1000
Dwell time (ms)	50	50	10
Gas Flow reaction cell gas ($mL\ min^{-1}$)	He / 1.5	NH_3 / 0.7	-
RPa	0	0	0
RPq	0.25	0.8	0.25
Signal acquisition	Continuous	Continuous	TRA
Nuclides monitored	$^{63}Cu^+$,$^{65}Cu^+$	$^{63}Cu^+$,$^{65}Cu^+$	$^{60}Ni^+$,$^{63}Cu^+$,$^{65}Cu^+$

Table 2. Instrumental conditions for Cu isotopic analysis of serum using direct injection to MC-ICP-MS.

Desolvating Nebulizer System ARIDUS3	
N$_2$ Gas (mL min^{-1})	0 [a]
Ar Gas (L min^{-1})	8
Chamber Temperature (°C)	140
Membrane temperature (°C)	160
Nu MC-ICP-MS 1700	
RF Power (W)	1300
Instrument resolution	Low
Integration time (s)	0.5
Nebuliser Pressure (Psi)	35.10 – 35.30 [b]
Auxiliary gas (L min^{-1})	0.8
Coolant gas (L min^{-1})	13

Faraday cup configuration																					
Collector	H9	H8	H7	H6	H5	H4	H3	H2	H1	Ax	L1	L2	L3	IC0	IC1	IC2	L4	IC3	L5	IC4	L6
m/z	66.6	66		65			64			63			62						60		

[a]The Aridus3 was not connected to any additional nitrogen flow, as in our experience this does not improve the sensitivity for the Nu

1700 instrument.

[b]: optimized daily

Table 3. Results obtained by the two laboratories for analysis of serum RM. Results
are expressed as $\bar{x}$ ± U, where U= (t.s)/√N for a 95 % confidence interval (N=3) and $\bar{x}$
is the mean value, t is the t-value, and s is the standard deviation.

	Laboratory in Pau			Laboratory in Zaragoza			Serum Reference material
	Total Cu (µg L^{-1})	CuEXC (µg L^{-1})	REC (%)	Total Cu (µg L^{-1})	CuEXC (µg L^{-1})	REC (%)	Total Cu (µg L^{-1})
RM L-1	1141 ± 20	362 ± 35	32 ± 3	1258 ± 37	424 ± 10	33 ± 1	1066 ± 215
RM L-2	1901 ± 61	1260 ± 51	66 ± 3	2009 ± 90	1419 ± 109	71 ± 3	1925 ± 387

Table 4. Results obtained with the μ-injection method through the analysis of the 3 μL of NIST SRM 3114 after the two consecutive procedures at different concentrations using the LRS method for the isotopic calculation. Cu recovery was 100 ± 3 %.

Cu concentration (μg L^{-1})	δ^{65}Cu "bracketing", ± 2SD (‰), (N=10)
250	-0.18 ± 0.24
500	-0.05 ± 0.26
1000	0.11 ± 0.19
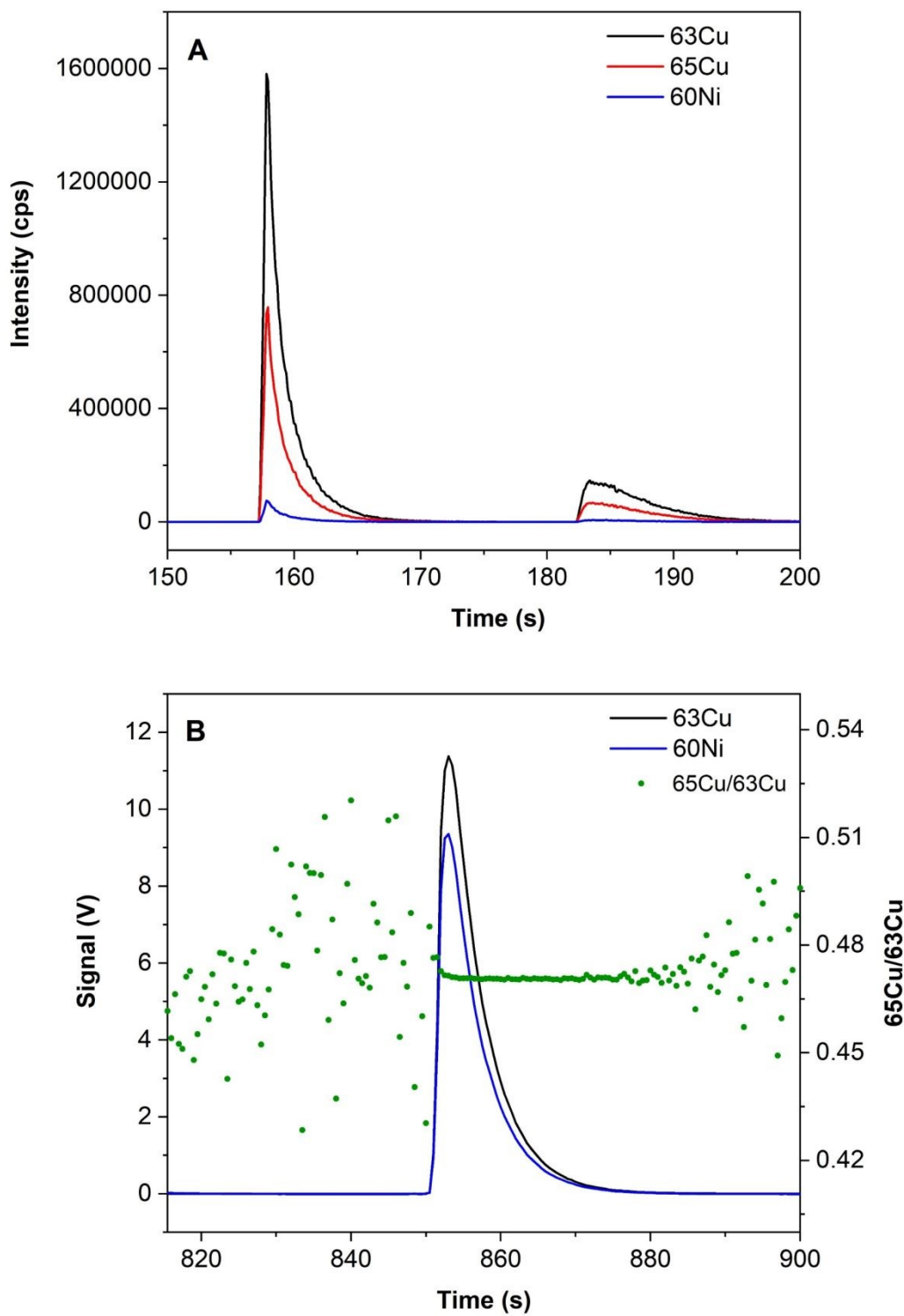
Figure 1. TRA Signal profiles obtained in a) Total Cu determination (1 µL) and b) Cu isotopic analysis (3 µL).
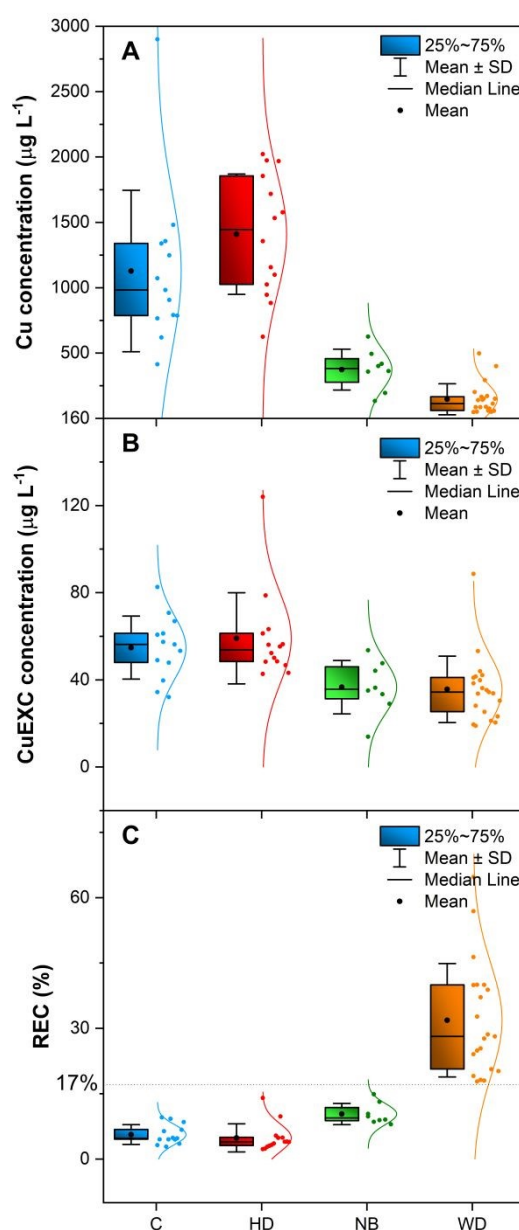
Figure 2. Results obtained from serum samples analyses for the four different groups considered in this study of: a) Total Cu concentration after Cu isolation step with Cu specific resin (Triskem); b) CuEXC concentration; and c) REC. C denotes controls, HD denotes patients with liver disease, NB denotes newborns and WD denotes patients with Wilson's disease. In addition to presenting the mean, the median, the standard deviation and the quartiles in the box plot, every dot beside the box represents each

individual measurement, and the peak-shape line represents the normal distribution fit

of such data.
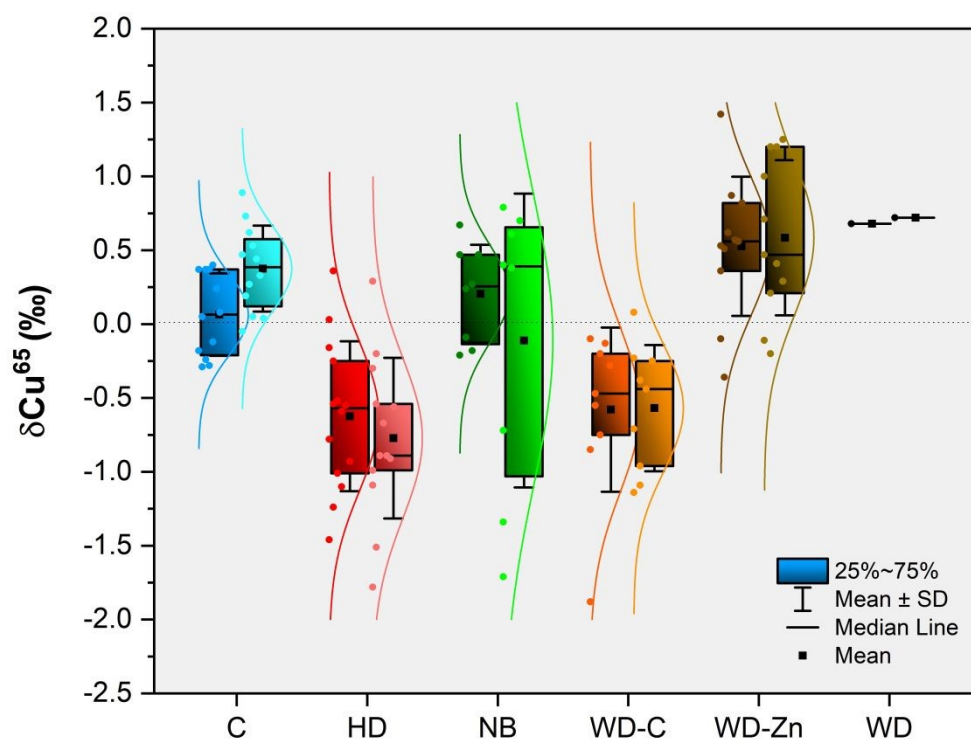
Figure 3. Boxplot obtained for Cu isotopic composition, expressed as $\delta^{65}Cu$ (‰) of both fractions of serum samples analyzed (Bulk serum, darker colors, and CuEXC fraction, lighter colors) for the six different groups considered in this investigation.
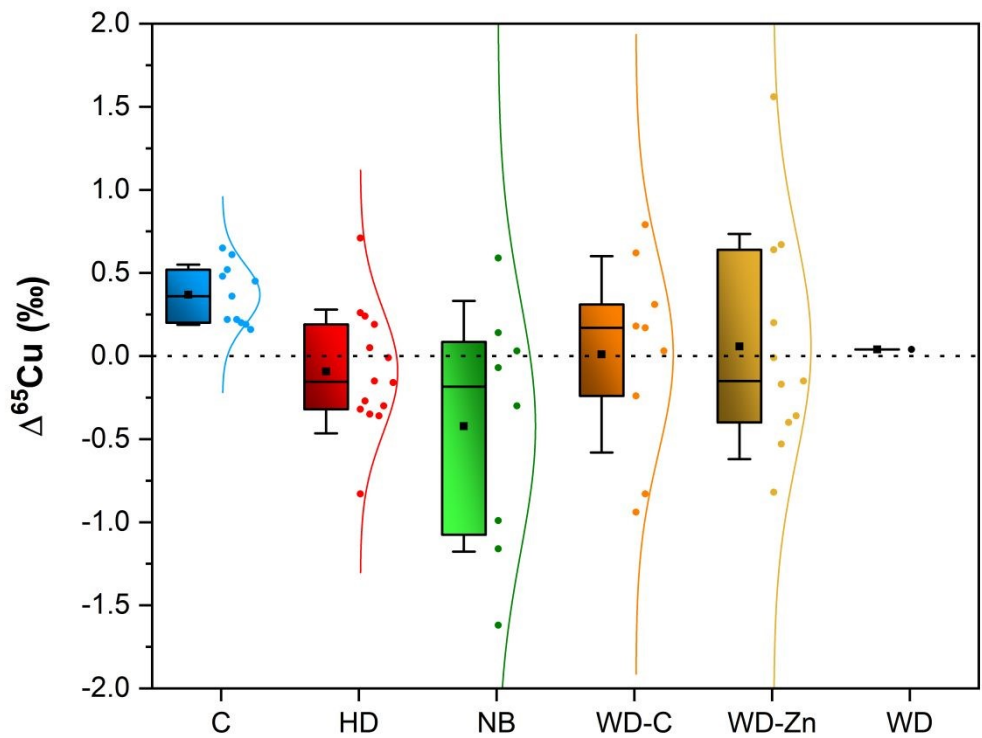
Figure 4. Boxplot obtained for $\Delta^{65}Cu$ (‰) for the six different groups considered in the current study.
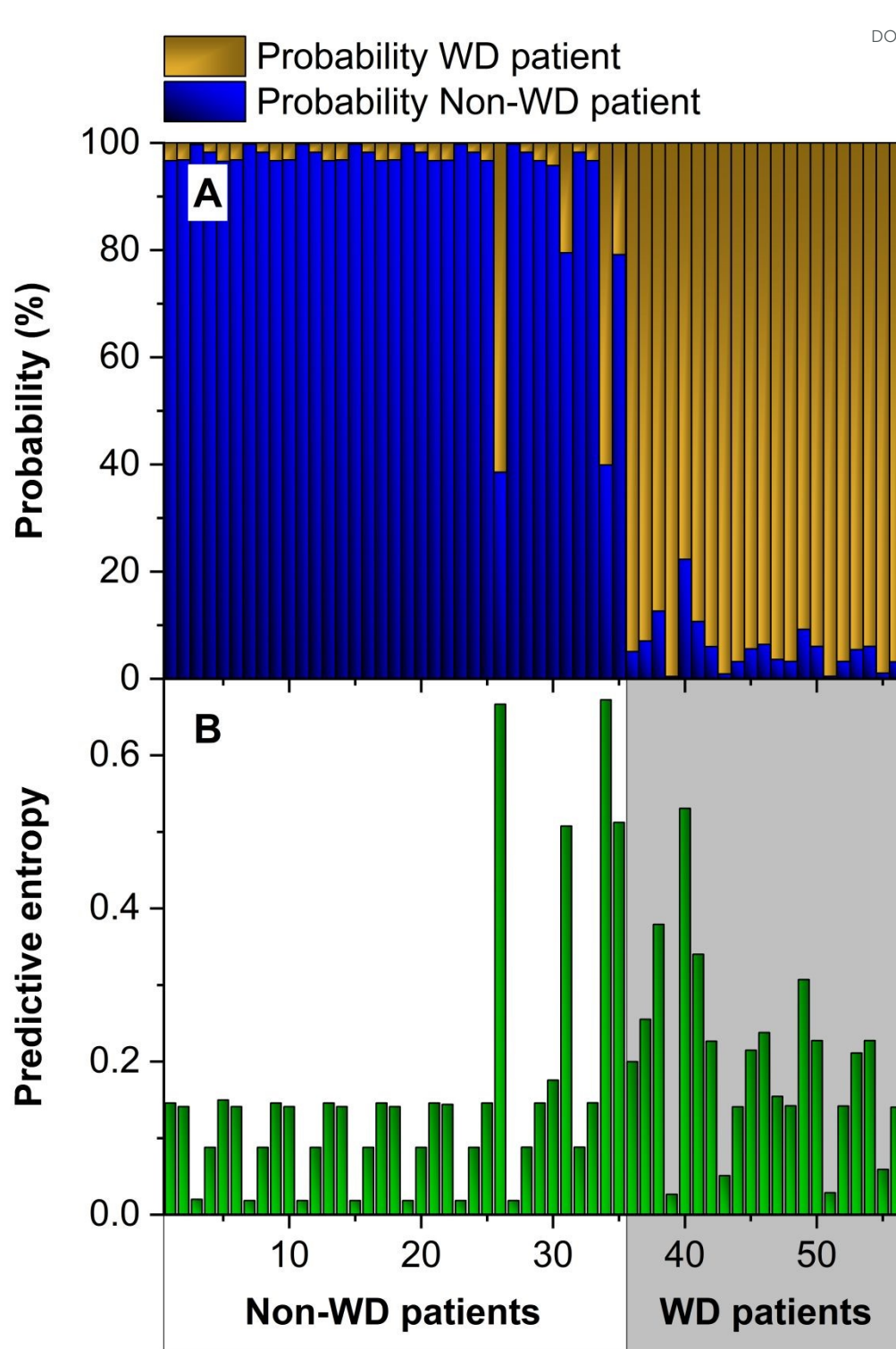
Figure 5. (A) Probability map and (B) predictive entropy obtained with the ensemble of 50 ANNs using the values of Cu total and CuEXC concentrations as inputs.
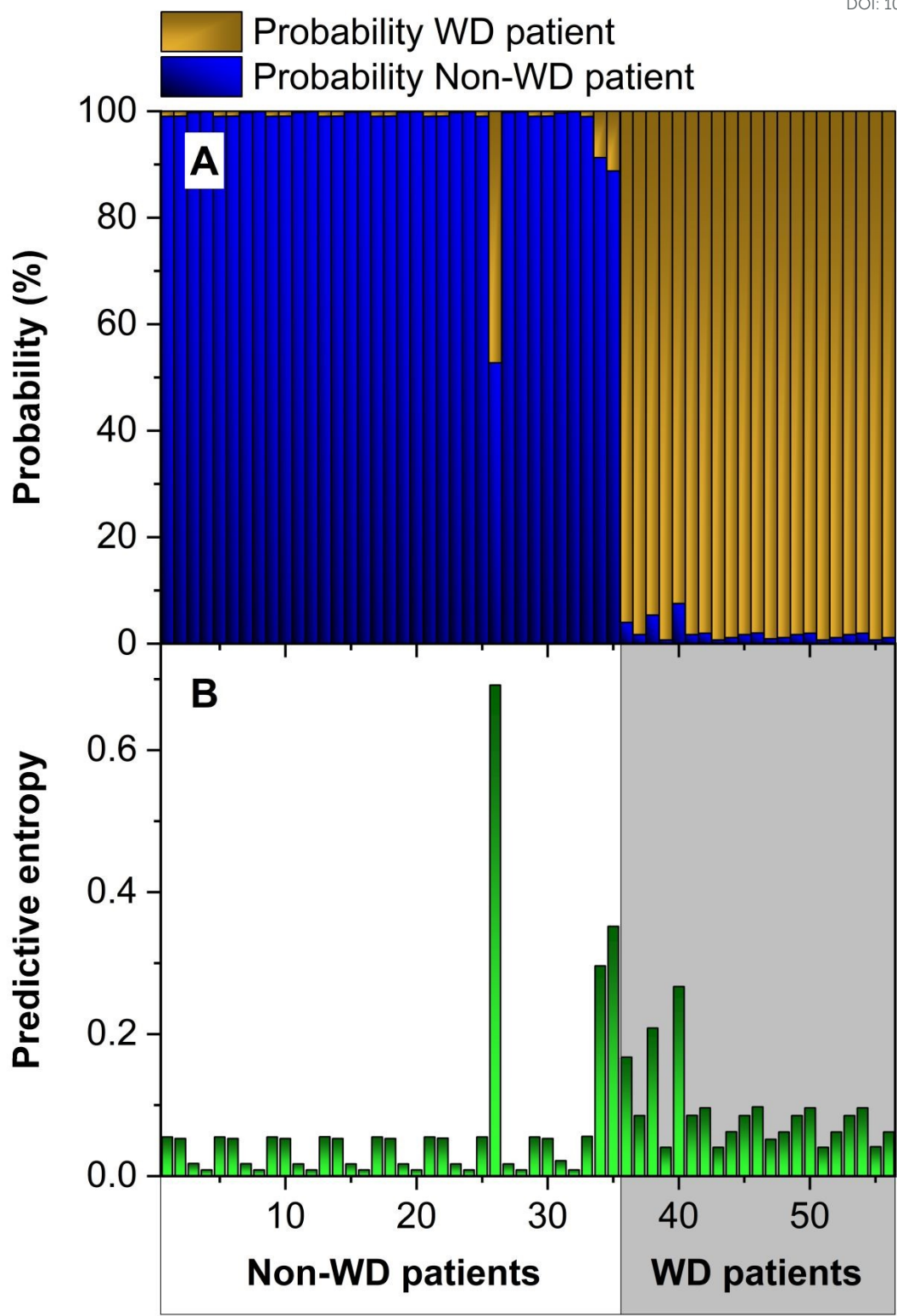
Figure 6. (A) Probability map and (B) predictive entropy obtained with the ensemble of 50 ANNs using the values of Cu total concentration, CuEXC concentration and REC as inputs.
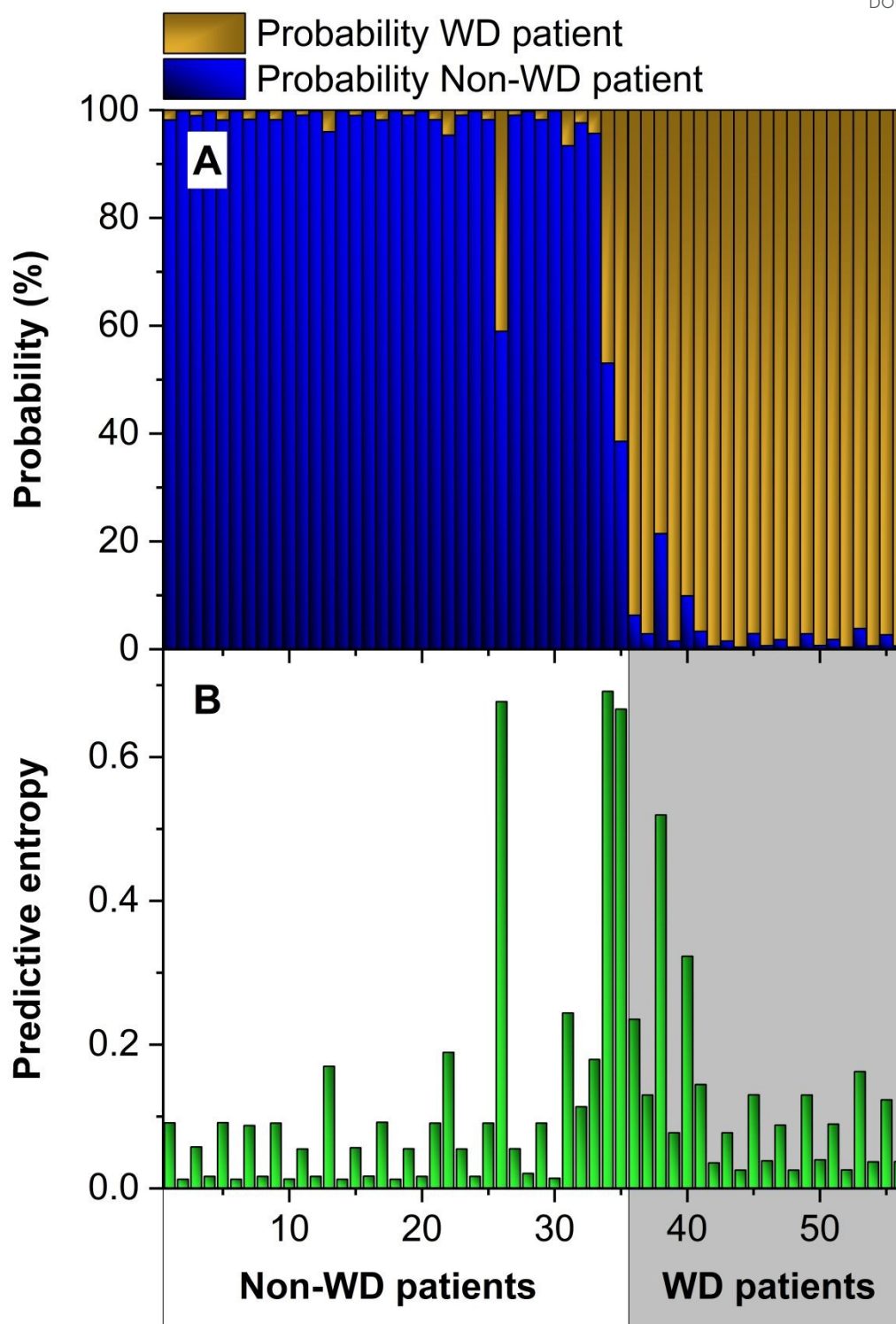
Figure 7. (A) Probability map and (B) predictive entropy obtained with the ensemble of 50 ANNs using the values of Cu concentration, CuEXC, REC, $\delta^{65}Cu_{bulk\ serum}$ and $\delta^{65}Cu_{EXC}$ as inputs.