



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV FYZIKÁLNÍHO INŽENÝRSTVÍ

INSTITUTE OF PHYSICAL ENGINEERING

**KOREKCE KALIBRAČNÍCH MODELŮ
SPEKTROSKOPIE LASEREM INDUKOVANÉHO
PLAZMATU PŘI ZMĚNĚ ABLAČNÍCH ENERGIÍ**

CORRECTION OF LASER-INDUCED BREAKDOWN SPECTROSCOPY CALIBRATION
MODELS FOR CHANGING ABLATION ENERGIES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

TOMÁŠ DVOŘÁK

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. ERIK KÉPEŠ

BRNO 2023

Abstrakt

Tato bakalařská práce se zabývá zkoumáním transfer learningu jakožto potenciální metody korekce kalibračních modelů na bázi vícevrstvných perceptronových neuronových sítí v rámci spektroskopie laserem indukovaného plazmatu v důsledku změny ablačních energií. Byly navrženy a natrénovány MLP kalibrační modely pro čtyři prvky (chrom, nikl, molybden a mangan) a šest různých ablačních energií (40, 50, 60, 70, 80 a 90 mJ). Modely původně natrénované na ablační energii 50 mJ byly poté použity k predikci koncentrací ze spekter měřených na různých ablačních energiích. Následné rozdíly v měření byly řešeny aplikací transfer learningu. Výsledky naznačují, že transfer learning by mohl sloužit jako validní metoda pro korekci nepřesností vznikajících v důsledku rozdílů v ablační energii, dosahující srovnatelných výsledků s modely trénovanými od základů za zlomek času a s výrazně nižší výpočetní náročností. Tato studie však nedokázala jednoznačně prokázat konzistentní zlepšení výkonu neuronových sítí pomocí transfer learningu v kontextu LIBS. Pravděpodobně přispívajícími faktory jsou nedostatečná optimalizace použitých neuronových sítí, omezená komplexita experimentálního datasetu nebo kombinace obou. Tento výzkum navazuje na a rozšiřuje literaturu a nabízí hlubší pochopení možností a omezení transfer learningu v kontextu LIBS.

Summary

This bachelor thesis investigates transfer learning as a potential method for correcting calibration models based on multilayer perceptron neural networks in laser-induced breakdown spectroscopy due to changes in ablation energies. MLP calibration models for four elements (chromium, nickel, molybdenum and manganese) and six different ablation energies (40, 50, 60, 70, 80 and 90 mJ) were designed and trained. Models initially trained at an ablation energy of 50 mJ were then used to predict concentrations from spectra measured at the different ablation energies. Subsequent differences in the measurements were resolved by applying transfer learning. The results suggest that transfer learning could serve as a valid method for correcting inaccuracies due to differences in ablation energy, achieving comparable results to models trained from scratch in a fraction of the time and at significantly lower computational cost. However, this study could not clearly demonstrate consistent improvements in neural network performance using transfer learning in the context of LIBS. Likely contributing factors are the lack of optimization of the neural networks used, the limited complexity of the experimental dataset, or a combination of both. This research builds on and extends the literature and offers a deeper understanding of the capabilities and limitations of transfer learning in the context of LIBS.

Klíčová slova

Spektroskopie laserem buzeného plazmatu, Transfer Learning, Fine-Tuning, MLP, Ablací energie

Keywords

Laser Induced Breakdown Spectroscopy, Transfer Learning, Fine-Tuning, MLP, Ablation Energy

DVOŘÁK, T. *Korekce kalibračních modelů spektroskopie laserem indukovaného plazmatu při změně ablačních energií*. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2023. 34 s. Vedoucí Ing. Erik Képeš.

Prohlašuji, že jsem bakalářskou práci zpracoval sám. Veškeré prameny a zdroje informací, které jsem použil k sepsání této práce, jsou citovány a uvedeny v seznamu použitých pramenů a literatury.

Tomáš Dvořák

Obsah

1	Úvod	2
2	Teorie	4
2.1	Spektroskopie laserem buzeného plazmatu	4
2.2	Kalibrace, regresní modely a hodnotící kritéria	4
2.3	Neuronové sítě (ANN)	6
2.4	Preprocessing dat pro ANN	9
2.5	Optimalizace ANN architektury	10
2.5.1	Hyperparametry ANN	11
2.5.2	Hodnocení výkonu ANN	11
2.5.3	Jevy zhoršující výkon ANN	12
2.5.4	Metody optimalizace ANN	13
2.6	Transfer learning	13
3	Praktická část	14
3.1	Cíle experimentu	14
3.2	Popis dat	14
3.3	Stanovení baseline přesnosti lineární regrese	15
3.4	Stavba a optimalizace architektury neuronové sítě	15
3.4.1	Preprocessing vstupních dat	15
3.4.2	Optimalizace ANN architektury	15
3.5	Transfer learning	18
4	Výsledky a diskuze	19
4.1	Lineární regrese	19
4.2	Stavba a optimalizace ANN	23
4.2.1	Optimalizace algoritmem hyperband	23
4.2.2	Manuální optimalizace	24
4.3	Transfer learning	25
5	Závěr	30

Kapitola 1

Úvod

Spektroskopie laserem buzeného plasmatu (Laser-Induced Breakdown Spectroscopy - LIBS) se v posledních letech stala významnou analytickou technikou v oblasti materiálové vědy, chemie a environmentální vědy [1]. Je schopná rapidní, nedestruktivní, multidimenzionální analýzy s minimální nutností přípravy vzorků. [2]

Princip LIBS spočívá v interakci vysokoenergetického laserového pulsu se vzorkem. Na povrchu analyzované oblasti dojde k ablací malého množství materiálu a jeho rozložení na plasma, tedy silně ionizovaný a excitovaný plynný stav. Plasma při své deexcitaci emituje charakteristické záření. Emitované záření se změří spektrometrem a jeho intenzita a tvar se analyzují za účelem určení chemického složení materiálů.

LIBS poskytuje hned několik výhod oproti ostatním běžným metodám chemické analýzy, jakými jsou například Ramanova nebo hmotnostní spektroskopie. Umožňuje nedestruktivní analýzu s minimální nutností přípravy vzorků [3], generuje multidimenzionální informace o prostorovém rozložení koncentrace měřených prvků [4], analýza pomocí LIBS je rychlá a může poskytnout výsledky v reálném čase, což otevírá dveře aplikacím v řízení a monitoringu procesů. LIBS má vysokou citlivost a dokáže detekovat stopové prvky. Pro analýzu pomocí LIBS je potřeba jen malé množství materiálu (řády μg) [5], což je užitečné pro analýzu malých vzorků a vzácných materiálů. Metoda je všestranná, je možné ji použít na širokou škálu materiálů včetně kapalin, plynů a organických materiálů [6]. LIBS laboratoře jsou malé a mohou být navrženy jako přenosné pro on-site analýzu. Jednou takovouto laboratoř je vybaven Mars Rover Curiosity [7].

Významným nedostatkem LIBS je potřeba kalibrace každé aparatury

zvášť pro každý typ měřeného vorku. Kalibrací přístroje se v tomto kontextu myslí stanovení vztahu mezi emisním spektrem plazmatu a elementárním složením vzorku. Tento proces obvykle zahrnuje analýzu vzorků s předem známým elementárním složením a použití těchto dat k vytvoření kalibrační křivky, matice nebo regresní neuronové sítě [8].

Hlavním důvodem potřeby kalibrace je přítomnost matričních jevů, což jsou jevy vyplývající z malých rozdílů v nehomogenitách složení měřených vzorků, a tedy rozdílných fyzikálních vlastnostech, jako je například tepelná vodivost, tepelná kapacita a optické vlastnosti. Tyto rozdíly mohou výrazně ovlivnit formaci plazmatu a přenos energie, a tím i tvar výsledného emisního spektra. Matriční jevy také násobí rozdíly mezi měřícími aparaturami. Ty se totiž liší ve faktorech s výrazným vlivem na tvorbu plazmatu, jako je energie laseru nebo délka trvání impulsu.

Přestože pro kalibraci LIBS aparatur zpravidla postačují vícerozměrné nelineární regresní modely (PLSR, SVR), vhodnou alternativou se v poslední době ukazují být neuronové sítě, převážně díky své schopnosti přizpůsobit se větším a komplexnějším datasetům. Svědčí o tom mimo jiné rostoucí množství publikací, které na toto téma vzniká [9]. Jednou ze zajímavých výhod neuronových sítí je jejich schopnost adaptovat se na nové, podobné datasety za pomoci transfer learningu - jevu, kdy si neuronová síť při tréninku ponechá naučené vzorce matrice, a pouze se je naučí aplikovat na novou, podobnou matici. Tato vlastnost má potenciál výrazně ušetřit čas v případě kalibrace nových aparatur na vzorcích, pro které již existuje kalibrační neuronová síť. V této práci se proto snažím prozkoumat validitu transfer learningu a zjistit jeho potenciální výhody jakožto metody rekaliibrace měřící aparatury pro různé ablační energie.

Kapitola 2

Teorie

2.1 Spektroskopie laserem buzeného plasmatu

LIBS funguje na principu analýzy charakteristických spektrálních čar mikroplasmatu vytvořného laserovou ablací na povrchu zkoumaného materiálu. K ablací je využíváno vysoce výkonného pulzního laseru, zpravidla typu Nd:YAG, o energii pulzu v rozpětí jednotek až stovek mJ. Laserový puls je soustavou čoček a zrcadel soustředěn na malou oblast povrchu vzorku, typicky desítky až stovky mikrometrů v průměru, která je vlivem soustředěné energie ablatována na mikroplasma. Formované plasma projde fázemi expanze a následného collapse. V collapse fázi jsou deexcitujícími kladnými ionty emitovány fotony charakteristických vlnových délek. Ty jsou sbírány soustavou čoček a soustředěny do optického vlákna, kterým jsou přivedeny na difrakční mřížku. Rozptýlené světlo tvoří spektrum, které je detekováno snímači CCD (Charge-Coupled Device) nebo CMOS (Complementary Metal-Oxide Semiconductor) [10].

2.2 Kalibrace, regresní modely a hodnotící kritéria

Regrese je statistická metoda používaná k modelování vztahu mezi závislou proměnnou a jednou nebo více nezávislými proměnnými. Regresní modely jsou matematické reprezentace těchto vztahů, umožňující závislosti mezi těmito proměnnými zkoumat a kvantifikovat. Cílem regresních modelů je najít nejlepší vztah mezi proměnnými tak, aby závislá proměnná mohla být předpovězena na základě hodnot nezávislých proměnných.

Predikční schopnosti regresních modelů se posuzují na základě hodnotících kritérií. Zpravidla se používají metriky směrodatná odchylka (Root Mean Square Error - RMSE) a absolutní odchylka (Mean Absolute Error - MAE). Jejich matematická vyjádření jsou následující [11]:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.1)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.2)$$

Přestože obě kritéria reprezentují rozdíl mezi předpovídanými a skutečnými hodnotami, RMSE silněji penalizuje odlehlé hodnoty a hrubě chybné předpovědi regresního modelu, a tedy lépe reflektuje jeho robustnost a výkon na co nejširším spektru vzorků. Máme-li však k dispozici více modelů, nejlepší přehled o jejich výkonu můžeme získat porovnáním obou metrik napříč jejich predikcemi na stejném datasetu.

Jako primárním hodnotícím kritériem pro posouzení přesnosti regresních křivek a neuronových sítí v této práci se řídím RMSE. Pro orientaci však také uvádím i MAE.

Jednoprvková lineární regrese je nejjednoduším druhem regresního modelu. Ve své základní podobě se jedná o fitování přímky jednorozměrným datasetem. Rovnice přímky je:

$$y = b_0 + b_1 x \quad (2.3)$$

Kde y je závislá proměnná, x je nezávislá proměnná, b_0 je průsečík s osou y a b_1 je směrnice přímky (poměr, ve kterém se y mění vzhledem k x). Hodnoty b_0 a b_1 jsou určeny metodou nejmenších čtverců tak, aby minimalizovaly RMSE mezi předpovězenými a skutečnými hodnotami y . Lineární regrese může být zobecněna na libovolné množství nezávislých proměnných, v takovém případě se jedná o víceprvkovou (nebo-li multivarietní či multidimenzionální) regresi a její tvar je:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_n x_n \quad (2.4)$$

Kde x_1, x_2, \dots, x_n jsou nezávislé proměnné a b_1, b_2, \dots, b_n jsou k nim přidružené regresní koeficienty. Výhodami lineární regrese jsou především jednoduchost výpočtu a následné interpretace a grafického zobrazení výsledků. Z těchto důvodů je lineární regrese také jednou z nejběžněji používaných statistických metod. Silným nedostatkem lineární regrese je její omezení na modelování lineárního vztahu modelovanou závislou proměnnou a vzájemně nezávislými

nezávislými proměnnými. Chceme-li takové vztahy modelovat, musíme sáhnout po komplexnějších statistických modelech, jako jsou například neuronové sítě.

V případě mého experimentu jsem použil jednorozměrnou lineární regresi. Jako závislou proměnnou jsem pro každý prvek určil jeho koncentraci a jako nezávislou proměnnou rozdíl mezi minimální a maximální hodnotou jedné jeho významné spektrální čáry.

2.3 Neuronové sítě (ANN)

Neuronová síť je druh algoritmu učení s učitelem (supervised learning). Skládá se z propojených uzlů nebo neuronů, které společně pracují na zpracování a analýze komplexních dat.

Neuronová síť pracuje tak, že přijímá vstupní hodnoty a postupně je předává skrze sérii propojených vrstev neuronů. Každý neuron na dané vrstvě aplikuje na váženou sumu vstupních hodnot z předchozí vrstvy matematickou funkci a produkuje výstup, který je poté předán další vrstvě. Výstup poslední vrstvy neuronů je potom celkovým výstupem neuronové sítě. Matematické funkce používané neurony se nazývají aktivační funkce.

Během procesu trénování neuronové sítě jsou váhy vstupních dat předávaných mezi jednotlivými neurony upravovány supervizovanými učícími algoritmy, jako je například SGD (Stochastic Gradient Descent). Tyto algoritmy fungují tak, že nechají síť dělat predikce na trénovacích datech se známou hodnotou predikované veličiny. Sítí předpovězená hodnota je pak porovnána se skutečnou hodnotou a váhy mezi každými dvěma neurony jsou upraveny tak, aby se zmenšil jejich rozdíl. Jakmile je neuronová síť natrénovaná, může být použita k predikci neznámých dat. Tento proces se nazývá učení s učitelem.

Nejjednodušší formou neuronové sítě je tzv. perceptron. Perceptron pracuje tak, že vezme sadu vstupních hodnot, vynásobí každou vstupní hodnotu odpovídající vahou a poté tyto vážené vstupy sečte. Výsledek této sumace je přiveden k aktivační funkci, která produkuje konečný výstup perceptronu. V matematické reprezentaci funguje perceptron následovně. Mějme vektor vstupů $x = [x_0, x_1, x_2, \dots, x_n]$ a vektor vah $w = [b, w_1, w_2, \dots, w_n]$. Nechť $x_0 = 1$

a necht se váha b nazývá bias. Vážená suma z se spočítá jako:

$$z = b + w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (2.5)$$

Výsledek perceptronu se poté spočítá aplikací aktivační funkce na váženou sumu:

$$y = f(z) \quad (2.6)$$

Kde $f(z)$ je aktivační funkce. Při tréninku jsou váhy perceptronu upraveny na základě chyby výstupu. Při regresních úlohách se jako chyba zpravidla používá RMSE. Pro i -tý bod vstupního vektoru x_i , predikovaný výstup y a skutečný výstup t , se následně upraví váha w_i :

$$\Delta w_i = \alpha(t - y)x_i \quad (2.7)$$

Kde α je experimentálně zvolená rychlost učení (learning rate).

Aktivační funkce používaná perceptronem pro binární klasifikaci je obvykle jednoduchá prahová funkce, která produkuje výstup 1, pokud je suma vážených vstupů větší než určitý práh, a výstup 0 v opačném případě. Všimněme si, že pokud jako aktivační funkci použijeme přímku (jako je tomu běžně při regresních problémech), po dosažení vážené sumy (2.5) do rovnice aktivační funkce (2.6), dostáváme stejný vztah jako v případě vícerozměrné lineární regrese (2.4). Perceptron je tedy stejně jako lineární regrese nedostatečující pro modelování nelineárních vztahů. Je však základem komplexnějších neuronových sítí, které toho již schopné jsou. Jednou takovou sítí je například MLP (multi-layer perceptron).

MLP je typ neuronové sítě složený z několika vrstev plně propojených perceptronů. To znamená, že každý perceptron v jedné vrstvě posílá svůj výstup každému perceptronu v následující vrstvě. Vrstva, která přijímá vektor nezpracovaných vstupních dat, se nazývá vstupní (input layer). Perceptrony vstupní vrstvy posílají svůj výstup sérií skrytých vrstev (hidden layers), na jejímž konci se nachází výstupní vrstva (output layer), jejíž výstup je výstupem celé neuronové sítě.

Váhy MLP se při tréninku upravují algoritmem zpětného šíření chyby (anglicky backpropagation). Tento algoritmus nejprve diferencíálem vypočítá odchylku δ_j^L příslušící každému j -tému vstupu x_j výstupní vrstvy L , tím tedy i každé aktivaci poslední hluboké vrstvy:

$$\delta_j^L = J'_L \left(\sum_i x_{ij} w_{ij} \right) \cdot (t - y^L) \quad (2.8)$$

V tomto vztahu J_L je aktivační funkce výstupní vrstvy a $t - y^L$ je rozdíl mezi předpovězenou a skutečnou hodnotou.

Tento proces můžeme zopakovat pro každý neuron l -té hluboké vrstvy a tím zjistit, jak se na jeho odchylkách podílí výsledky aktivací předchozí vrstvy $l+1$. Součtem těchto odchylek přes celou vrstvu $l+1$ získáme celkovou chybu každého neuronu vrstvy l . Pro j -tý neuron libovolné vrstvy l a každý z celkem k neuronů vrstvy $l-1$ můžeme získat jeho chybu δ_j^l vztahem:

$$\delta_j^l = J'_l\left(\sum_i x_{ij} w_{ij}\right) \cdot \sum_k (w_{kj}^{l-1} \delta_k^{l-1}) \quad (2.9)$$

Váha mezi j -tým neuronem a i -tou aktivací y_i^{l-1} neuronu z předchozí vrstvy se poté upraví podle vztahu:

$$\Delta w_{ij}^l = \alpha \delta_j^l y_i^{l-1} \quad (2.10)$$

Nejpoužívanější aktivační funkcí pro neurony skrytých vrstev je ReLU (Rectified Linear Unit). Je definována následujícím způsobem:

$$f(x) = \max(0, x) \quad (2.11)$$

Hlavním důvodem za popularitou této funkce je její výpočetní nenáročnost podpořená možností sparse reprezentace nulových perceptronů. Její další silnou výhodou je schopnost přinášet do ANN nelinearitu, a přitom nepodléhat efektům vanishing a exploding gradientu. Vanishing a exploding gradient jsou negativní jevy, ke kterým dochází při zpětném šíření chyby, kdy první derivace některých nelineárních aktivačních funkcí, jakými je například sigmoid nebo hyperbolický tangens, v případě nízkých vstupních hodnot nabývá extrémně vysokých čísel, zatímco v případě vysokých vstupních hodnot se blíží nule.

Počet neuronů výstupní vrstvy a jejich aktivační funkce závisí na nátuře problému, který má neuronová síť řešit. U regresních problémů se zhora neohrazeným výstupem se zpravidla využívá jeden perceptron s lineární aktivační funkcí, při nebinární klasifikaci má výstupní vrstva tolik perceptronů, jako je tříd, mezi kterými neuronová síť rozlišuje. Jejich aktivační funkce jsou běžně binární funkce reprezentující pravděpodobnost, že zkoumaný vzorek náleží dané třídě, jako je například softmax:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (2.12)$$

Rovnice bere jako vstup váženou sumu vstupů pro j -tý neuron, z_j , a počítá pravděpodobnost j -té třídy jako funkci exponenciálních hodnot všech

výstupních neuronů. To zajistí, že součet výstupních pravděpodobností bude vždy 1.

Hlavní výhodou MLP oproti samotným perceptronům je jejich schopnost modelovat složité nelineární vztahy mezi vstupními a výstupními daty. To umožňuje generalizaci a řešení komplexních problémů s velkým množstvím nezávislých proměnných. Na druhou stranu, silným nedostatkem MLP je obtížnost pochopení, interpretace a vizualizace jimi učiněných závěrů. MLP se také často vyznačují vysokou dimenzionalitou, zejména v případech, kdy je k dispozici velké množství vstupních dat. Vzhledem k exponenciálnímu nárůstu počtu parametrů s rostoucím počtem neuronů je často nutné uplatnit techniky, jako je pruning, regularizace nebo redukce dimenzionality, aby se minimalizoval čas trénování na přijatelnou dobu. V některých případech může být také vhodné použít jinou architekturu ANN, která se dokáže efektivně zpracovávat vícedimenzionální vstupy. Příkladem takové architektury je konvoluční neuronová síť (CNN).

V CNN jsou vstupní data přivedena skrze sérii konvolučních vrstev, následovaných pooling vrstvami a poté připojena k jedné nebo více vrstvám plně propojených perceptronů. Základním prvkem konvoluční vrstvy je filtr - perceptron, který klouže po celé délce vektoru vstupních dat (podobně jako například funkce průměr při výpočtu klouzavého průměru), pro každý vstupní bod spočítá váženou sumu oblasti v jeho okolí a dosadí ji do aktivační funkce. Výstupní signál y_i filtru šířky m pro nějaký vstupní bod x_i můžeme poté s ohledem na vztahy (2.5) a (2.6) vyjádřit následujícím způsobem:

$$y_i = J((wx)_i + b) = J \left(\sum_{j=0}^{m-1} w_j x_{i-j} + b \right) \quad (2.13)$$

Kde w je vektor vah délky m , b je bias neuronu a J je aktivační funkce. Po každé konvoluční vrstvě zpravidla následuje pooling vrstva. Pooling vrstva je funkce, která bere vstupní hodnoty a vytváří redukovanou množinu aplikováním agregace nebo zmenšení (downsampling). Příkladem pooling algoritmu oblíbeného v CNN je max pooling, který bere maximální hodnotu z každé oblasti, která má být agregována.

2.4 Preprocessing dat pro ANN

Preprocessing jsou metody zpracování dat aplikované na trénovací a testovací data před jejich podáním neuronové síti. Zpravidla se jedná o metody čištění, redukce a transformace. Hlavním účelem preprocessingu bývá zlepšení

použitelnosti dat pro trénování, a tím snížení potřebné trénovací doby, dimenzionality modelu a zlepšení jeho predikčních schopností. Běžným procesem transformace dat používaným při tréninku neuronových sítí je standardizace. Při té se vstupní data přeškálují tak, aby byl v každém rozměru průměr 0 a standardní odchylka 1, což zabraňuje optimalizačnímu algoritmu upřednostňovat rozměry přirozeně obsahující vyšší hodnoty, a tím zpravidla vede k rychlejší konvergenci. Standardizaci lze vyjádřit následující rovnicí:

$$f(x) = \frac{x - \mu}{\sigma} \quad (2.14)$$

kde x jsou vstupní data, μ je průměr x , σ je standardní odchylka x a $f(x)$ jsou standardizovaná data. Alternativou ke standardizaci je normalizace. To je technika, která přeškáluje vstupní data na hodnoty od 0 do 1. To může být užitečné v případech, ve kterých jsou rozdíly ve škále mezi nezávislými proměnnými (vstupní data jsou například vzdálenost a čas). Tento přístup je velice citlivý na odlehlé hodnoty a nedoporučuje se používat jej na data-setech, kde se vyskytují. Normalizační funkce se obvykle vyjadřuje pomocí následující rovnice:

$$f(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2.15)$$

kde x jsou vstupní data a $f(x)$ jsou normalizovaná data.

Čištění datasetu se provádí za účelem odstranění šumu a pozadí v datasetu. Mezi nejčastěji používané metody při spektrální analýze patří rozklad na vlnky, který umožňuje rozložení signálu na různé frekvenční složky a tím eliminaci šumu obsaženého v těch nejvyšších vrstvách. Další častou metodou odstranění šumu u spektroskopických dat je průměrování několika spekter dohromady.

Cílem redukce datasetu je odstranění zbytečných nebo vysoce korelovaných nezávislých proměnných, což vede ke snížení počtu vstupních hodnot a tedy exponenciálnímu snížení komplexity a času trénování neuronové sítě. Jednou z metod redukce dimenzionality může být již výše zmíněná transformace na vlnky, při které mohou být nezávislé významné vlnky použity jako vstupní data pro neuronovou síť.

2.5 Optimalizace ANN architektury

Optimalizace architektury ANN neboli tuning zahrnuje hledání optimální kombinace hyperparametrů neuronové sítě pro dosažení co nejlepších predikčních schopností, případně dodatečně co nejkratšho času tréninku či predikce. Těchto cílů tuning dosahuje odstraněním některých negativních jevů

souvisejících se špatně nastavenými hyperparametry. Tuning ANN zahrnuje opakované trénování, testování a úpravu hyperparametrů sítě. Provádí se buď manuálně nebo algoritmicky.

2.5.1 Hyperparametry ANN

Hyperparametry jsou nastavení, která definují architekturu a trénink neuronové sítě. Tyto nastavení jsou obvykle stanovena výzkumníkem nebo vývojářem pracujícím na neuronové síti a mohou významně ovlivnit výkon sítě. Některé běžné hyperparametry neuronové sítě zahrnují:

1. Počet vrstev sítě. Neuronová síť s mnoha vrstvami se může naučit složitější vztahy, ale její trénink může být výpočetně náročnější. Hlubší neuronové sítě jsou také náchylné na jevy zvané vanishing a exploding gradient, ty se dají ale kompletně obejít použitím aktivační funkce ReLU.

2. Počet neuronů v každé vrstvě. Větší počet neuronů může zvýšit schopnost sítě naučit se složitější vztahy, ale může také zvýšit její výpočetní náročnost a riziko přetrénování.

3. Aktivační funkce. Různé aktivační funkce se mohou lépe hodit k charakterizaci rozdílných vztahů. V hlubokých vrstvách se ale zpravidla používají funkce ReLU (případně leaky ReLU), zejména díky svojí výpočetní nenáročnosti a řešení problémů s jevy vanishing a exploding gradientu.

4. Rychlost učení. Vyšší rychlost učení může vést k rychlejší konvergenci, ale může také způsobit oscilaci sítě kolem optimálního řešení a overfitting.

5. Optimalizační algoritmus. Optimalizační algoritmy zpravidla vycházejí z funkce gradientního sestupu (gradient descent).

6. Počet epoch. Počet epoch odkazuje na počet průchodů celého trénovacího datasetu. Vyšší počet epoch může zlepšit schopnost sítě naučit se složité vzory, ale zároveň může vést k overfittingu a zvýšení výpočetní náročnosti. Je důležité najít správnou rovnováhu mezi počtem epoch a dosažením optimálního výkonu sítě.

2.5.2 Hodnocení výkonu ANN

Výkon neuronové sítě se posuzuje na základě hodnotících kritérií. K vyhodnocování klasifikačních úloh jsou obvykle používány míry přesnosti, senzitivity, specifčnosti a F1 skóre, zatímco k vyhodnocování úloh kalibrace se jako u ostatních regresních úloh používají metriky RMSE a MAE, definované v sekci [2.2 \[12\]](#).

Chceme-li mít realistickou představu o výkonu modelu, je nezbytné měřit jeho hodnotící kritéria na odlišném datasetu, než na jakém byl trénován a

optimalizován. Z tohoto důvodu je před optimalizací běžnou praxí vyčlenit ze vstupního datasetu validační a testovací data. Na validačních datech je model testován mezi jednotlivými kroky optimalizace hyperparametrů, na testovacích datech jsou posuzovány jeho předpokládané predikční schopnosti po optimalizaci hyperparametrů. [13]

Může se také stát, že optimalizujeme hyperparametry neuronové sítě příliš dobře pro specifickou skupinu validačních dat. Populární způsob, jak tomuto předejít, se nazývá k -násobná cross-validace. Při té se trénovací dataset rozdělí na k skupin o přibližně stejném počtu vzorků tak, aby každá skupina co nejlépe reprezentovala variaci celé matice. Poté je neuronová síť trénována na vzorcích z $k - 1$ skupin a validována na vzorcích zbylé skupiny. Tento proces je opakován pro každou skupinu. Robustní neuronová síť by měla nést konzistentní výsledky napříč skupinami a hodnotícími kritérii [14].

2.5.3 Jevy zhoršující výkon ANN

Špatně nastavené hyperparametry neuronové sítě se obvykle projevují díky specifickým jevům, které zhoršují výkon sítě. Tyto jevy mají charakteristické projevy, a proto existují specifické metody pro jejich detekci a řešení.

Jedním z častých problémů je overfitting. Ten nastává, když se příliš komplexní a dlouho trénovaný model, případně model trénovaný na nedostatečném množství dat, začne přizpůsobovat specifitám a šumu v trénovacích datech a začne se odnaučovat obecné vztahy aplikovatelné na nová, dříve neviděná data. Overfitting vede ke zdánlivému zlepšení přesnosti na trénovacích datech, ale jejímu zhoršení na testovacích datech. Metody adresující overfitting se zpravidla zakládají na zvýšení komplexity [15].

Opačným jevem k overfittingu je underfitting. K tomu dochází, když je model příliš jednoduchý na to, aby v datech zachytil základní struktury. Underfitting se zpravidla projevuje špatným výkonem na trénovacích i validačních datech a model při tréninku často konverguje k aritmetickému průměru. Underfitting lze řešit zvýšením složitosti modelu, přidáním více parametrů nebo zvýšením množství trénovacích dat [15].

Dalším problémem při tuningu ANN je pomalá konvergence, při které neuronová síť při učení příliš dlouho konverguje k optimálnímu řešení. To může být způsobeno faktory, jako je nevhodná rychlost učení, špatná inicializace parametrů modelu nebo neoptimální optimalizační algoritmy. Pomalá konvergence může také vést k zaseknutí optimalizačního procesu na lokálním minimu, čemuž se dá ale předcházet zvýšením stochasticity modelu [16].

2.5.4 Metody optimalizace ANN

Existuje několik přístupů k optimalizaci architektury neuronové sítě. Tradičním přístupem je optimalizace manuální. Při té jsou architektem sítě po jednom upravovány její hyperparametry tak, aby bylo dosaženo optimálního výkonu. Při úpravě hyperparametrů se vývojář zpravidla řídí grafy vývoje ztráty a validační stráty při trénování sítě, ve kterých hledá projevy nepřesného nastavení. [17]

Alternativou k manuální optimalizaci je optimalizace algoritmická. Trénovací algoritmy umožňují architektovi nastavit konfigurace hyperparametrů, které budou po jedné natrénovány a otestovány. Nejjednoduššími optimalizačními algoritmy jsou grid search a random search. Grid search testuje všechny kombinace hodnot nastavených hyperparametrů, random search testuje předem stanovené množství náhodně vybraných kombinací. [18]

Hyperband je optimalizační algoritmus vylepšující random search o metody jako je adaptivní alokace zdrojů a early stopping. Algoritmus pracuje tak, že v průběhu několika kol, každého sestávajícího z progresivně vyššího množství epoch, vybírá z náhodně vygenerované konfigurace hyperparametrů ty slibné, zatímco ty méně slibné odstraňuje [19]. Algoritmus se ukázal jako velmi silný nástroj na optimalizaci výpočetně náročných neuronových sítí, pro méně komplexní sítě však projevil tendence vytvářet přeoptimalizované modely s nízkou schopností generalizovat [20].

2.6 Transfer learning

Transfer learning je technika, při které se při trénování neuronové sítě místo nového modelu s nulovými vstupními hodnotami použije již hotový model natrénovaný a optimalizovaný na podobné, související úloze [21].

Transfer learning je založen na ideji, že modely hlubokého učení, naučené rozpoznávat určité rysy ve vstupních datech, tuto znalost při přetrénování neztratí, nýbrž ji naučí aplikovat na nový typ dostatečně podobných vstupních dat [22].

Transfer learning je užitečný zejména v případech, kdy je nedostatek trénovacích dat pro konkrétní úlohu nebo kdy máme nedostatečné výpočetní zdroje k natrénování modelu od začátku [23].

Kapitola 3

Praktická část

3.1 Cíle experimentu

Cílem mého experimentu bylo prozkoumat vliv, jaký má změna ablační energie na kalibrační modely ve formě neuronových sítí, a prozkoumat validitu transfer learningu jakožto potenciální metody korekce tohoto vlivu. K experimentu jsem použil dataset 53 kalibrovaných vzorků, měřených na šesti různých energiích ablačního laseru - 40, 50, 60, 70, 80 a 90 mJ. Nejprve jsem proto porovnával výkon a počet epoch potřebných ke konvergenci přeučené neuronové sítě oproti náhodně inicializovaným neuronovým sítím o stejných hyperparametrech vyjma rychlosti učení, která byla optimalizovaná na jedno desetinné místo. V druhé části experimentu jsem sledoval závislost výkonu plně konvergované přetrénované neuronové sítě na některých jejích hyperparametrech a parametrech testování, specificky learning rate, batch size, počtu vzorků a počtu měření každého vzorku.

3.2 Popis dat

Experiment jsem prováděl na neupravené verzi dat poskytnutých v soutěži LIBS 2022 quantification contest [24]. Dataset obsahuje celkem 53 vzorků příbuzných slitin kovů. Každý vzorek byl měřen 200 krát pro každou z energií laseru 40, 50, 60, 70, 80, 90 mJ, což umožnilo vytvoření rozsáhlého množství spekter s různými signálními úrovněmi a šumem. Měřené spektrum zahrnuje rozsah vlnových délek od 240 až 900 nm. Referenční měření bylo provedeno na hmotnostním spektrometru Q4 Tasman (Bruker, GE) Odborem slévárenství na Ústavu strojírenské technologie FSI VUT. Celkově tento dataset poskytuje významné výzvy pro kvantitativní analýzu spektrálních dat a je vhodným prostředím pro testování a porovnání různých metod a algoritmů.

3.3 Stanovení baseline přesnosti lineární regrese

Pro získání přehledu o kvalitě dat měřených na každé ablační energii jsem jako první krok experimentu sestavil lineární regresní křivky energií na základě spektrálních čar Cr II 375.8 nm, Mo I 715.4 nm, Ni II 361.9 nm, Mn I 294.9 nm. Podle těchto kalibračních křivek jsem detekoval a eliminoval odlehle hodnoty, jejichž přítomnost v datasetu by vedla ke zkreslení až úplnému znehodnocení výsledků neuronových sítí. Při určování hodnotících kritérií jsem rozdělil data na testovací a trénovací stejným způsobem, jako tomu bylo při soutěži.

3.4 Stavba a optimalizace architektury neuronové sítě

3.4.1 Preprocessing vstupních dat

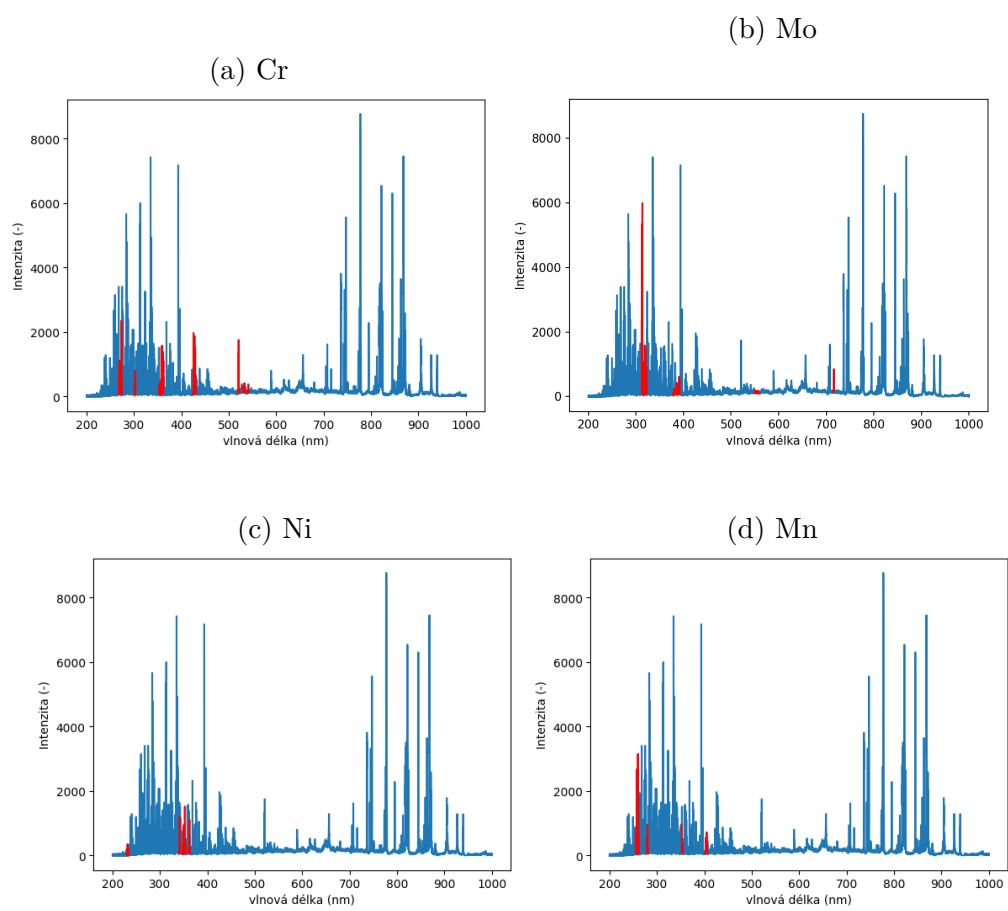
Při preprocessingu vstupních dat pro neuronovou síť jsem se rozhodl řídit zjednoušenou verzí postupu použitého vítězným týmem soutěže LIBS 2022 Quantification Contest [25]. Ze spekter jsem vyřezal oblasti, na kterých by se podle databází NIST a Atomtrace měly vyskytovat silné spektrální čáry každého prvku [?]. Tyto oblasti jsem dále pouze standardizoval na základě průměru a směrodatné odchylky trénovacích dat 2.14, než jsem je použil jako vstupní data pro MLP.

Vzorky použité při soutěži jako testovací jsem dále nepoužíval v rámci tréninku a validace. Z množiny trénovacích dat jsem manuálně vyseletoval šest vzorků jako validační tak, aby oba datasety měly přibližně stejný rozptyl a směrodatnou odchylku.

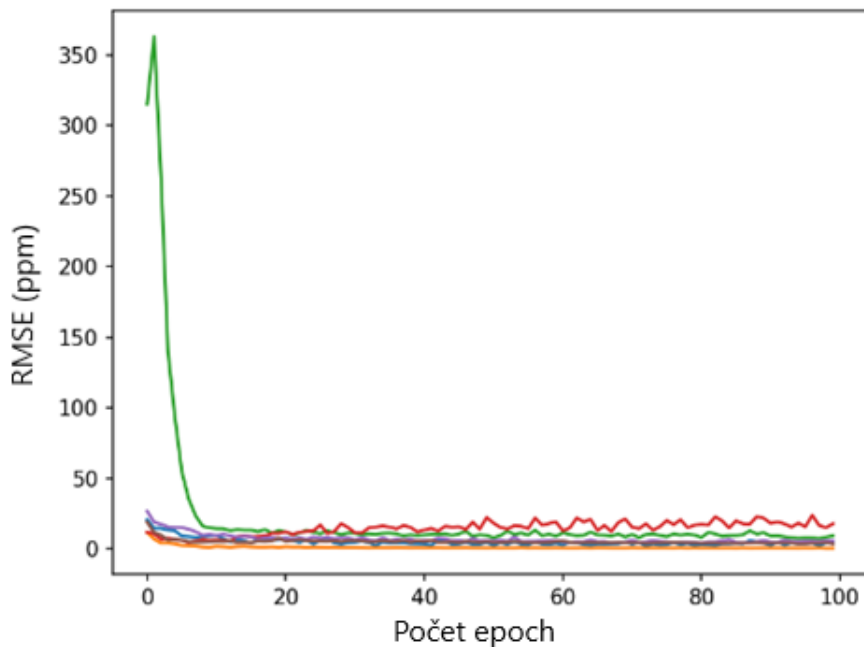
3.4.2 Optimalizace ANN architektury

Optimalizačním algoritmem hyperband jsem vygeneroval 10 modelů optimalizovaných pro nejnižší validační RMSE.

Na každém modelu jsem provedl 6-násobnou cross-validaci na trénovacích datech rozřazených do 6 skupin podle koncentrace každého prvku. Modely bez výjimky drasticky divergovaly pro jednu až dvě cross-validační skupiny, což je v souladu s nálezy výzkumníků [20]. Z grafů průběhu učení je možné



Obrázek 3.1: Izolované spektrální čáry jako vstupní data do ANN



Obrázek 3.2: Graf CV ztráty nejvýkonnější architektury generované algoritmem hyperband

vyvodit, že důvodem divergence může být náchylnost modelů k overtrainingu pro některé specifické subsety. Ke zvýšení konzistence tréninku jsem přidal dropout vrstvy jakožto regularizační metodu [26]. Vybral jsem model s nejnezjistitelnějším výkonem 3.2, zaokrouhlil počty neuronů v jeho hlubokých vrstvách a manuálně tunoval jeho dropout rate. Testoval jsem kombinace hodnot 0, 0.1, 0.2, 0.3, 0.5. Pro každou kombinaci jsem vyzkoušel několik hodnot learning rate z následujících možností, dokud jsem nedosáhl lokálního optima: 10^{-3} , $3 \cdot 10^{-3}$, 10^{-4} , $3 \cdot 10^{-4}$, 10^{-5} . Každých několik kroků jsem se na množině mocnin čísla 2 od 1 do 16 pokusil najít optimální batch size. Nakonec jsem skončil s následující neuronovou sítí:

Tato síť nesla průměrné RMSE pro cross-validační subsety (odkaz na výsledky) a na testovacích datech (odkaz na výsledky). Model jsem následně natrénoval na všech testovacích a validačních datech a validoval jej na datech testovacích.

Z vývoje validační ztráty jsem pozoroval, že neuronová síť nekonvergovala po běžně používaném množství epoch (100), ale až 300 epoch na to. Protože experimentuji s transfer learningem, je nezbytné pracovat s úplně

Tabulka 3.1: Neural Network Architecture

Layer	Units/Rate	Parameters
Input	1957	1957
Dense 1	150	293700
Dropout 1	0.5	0
Dense 2	150	22650
Dropout 2	0.5	0
Output	150	151
Learning rate:	0.00003	
Batch size:	1	

zkonvergovanou sítí, jinak nemohu při jejím přetrénování určit, zda ke konvergenci dochází v důsledku transfer learningu nebo pouze jejího trénování. Z tohoto důvodu jsem si dovolil dotrénovat sít dalších 300 epoch.

Model se stejnou architekturou vrstev a learning ratem optimalizovaným na jednu platnou cifru jsem natrénoval pro každou energii. K optimalizaci learning ratu jsem použil stejný cross-validační algoritmus. Tyto modely poslouží jako baseline pro srovnání s výkonem přetrénovaných neuronových sítí v pozdější části experimentu.

3.5 Transfer learning

Jako základ pro přetrénování jsem vybral neuronové sítě trénované na ablační energii 50 mJ. Tunoval jsem batch size, learning rate a dropout rate pro minimalizaci RMSE chromu na 90 mJ. Batch size jsem testoval na gridu mocnin 2 od 1 do 32, learning rate na logaritmickém gridu od 10^{-3} do 10^{-5} a dropout pro hodnoty 0, 0.3, 0.5 na obou vrstvách současně. Úprava learning ratu neovlivňovala výsledky tréninku do významné míry nezávisle na ostatních hyperparametrech, proto jsem se rozhodl použít stejnou hodnotu jako při trénování. Zvýšení batch size a odstranění dropout vrstev korelovalo pozitivně s výkonem přetrénovaných sítí. Experiment jsem zopakoval pro mangan a ablační energii 40 mJ se srovnatelnými výsledky. Ke konvergenci zpravidla docházelo během sedmé až osmé epochy. Na základě těchto měření jsem zvolil hyperparametry learning rate $3 \cdot 10^{-4}$, batch size 32 a dropout obou vrstev 0. Na těchto hyperparametrech jsem trénoval všechny modely po dobu 10 epoch a sledoval jejich konvergenci. Všechny modely do tohoto počtu epoch konvergovaly.

Kapitola 4

Výsledky a diskuze

4.1 Lineární regrese

Pro měřené prvky jsem sestrojil lineární regresní křivku na základě rozdílu maxima a minima následujících spektrálních čar: Cr II 375.8 nm, Mo I 715.4 nm, Ni II 361.9 nm, Mn I 294.9 nm. Spektrální čáry jsem určil na základě databáze Atomtrace[27]. Pro molybden jsem testoval 5 nejsilnějších spektrálních čar, žádná nevykazovala vyšší intenzitu, než je hladina šumu. Demonstrativně jsem vybral spektrální čáru Mo I 715.4 nm, protože byla v databázi uvedena jako nejsilnější. Zde se odkazuji na článek, kde byla ve stejných datech metodou analýzy důležitosti příznaků detekována jediná spektrální čára Mo II 281.6 nm [28], mě se tento výsledek však nepodařilo replikovat.

Na základě výsledků lineární regrese Cr pro ablační energii 80 mJ jsem vyselekoval data se směrodatnou odchylkou přesahující 3σ jako odlehlé hodnoty a odstranil je z trénovacího i testovacího datasetu. Identifikoval jsem tak vzorky 270121_5732PV, 200421_PN, 150921_ZP, 010721_PN z trénovacích dat a 011220_PN, 270121_5721PN z testovacích dat. Propozici, že se skutečně jedná o odlehlé hodnoty, je možné podpořit faktem, že tyto vzorky nebyly přítomny v datasetu soutěže LIBS 2022 quantification contest.

Z hodnotících kritérií pro první sadu regresních křivek je partneré, že lineární regrese bez odstranění odlehlých hodnot, případně bez aplikace jiné metody úpravy vzorků, nedostačuje jako kalibrační metoda pro žádný z prvků. To je indikováno nízkými až zápornými hodnotami R^2 . Nejhorší výsledky vykazoval molybden se všemi hodnotami R^2 zápornými. Nejlepší výsledky bylo možné sledovat u manganu na ablační energii 40 mJ, s R^2 0.41.

Po odstranění vzdálených hodnot došlo ke zlepšení výkonu napříč všemi prvky, nejvíce však pro prvky Cr a Ni, u nichž se nyní stala patrnou silná

lineární závislost. Po odstranění odlehlých hodnot došlo u všech prvků vyjma molybdenu ke změně vztahu mezi výkonem a ablační energií. Trend chromu a manganu se posunul od negativní korelace k neurčitě, zatímco trend niklu se přesunul od neurčitě k pozitivní. Nejpravděpodobnějším vysvětlením tohoto chování je přítomnost matričních jevů, které jsou charakteristické přínosem nelinearity pozitivně korelující s intenzitou silných spektrálních čar a negativně s intenzitou slabých spektrálních čar. Odlehlé hodnoty, jakožto píky s výrazně vyšší intenzitou, jimi byly ovlivněny do vyšší míry. Tato příručka popisuje pozitivní korelaci mezi výškou píky a projevem matričního jevu [29]. V této práci se diskutuje disproporcionální vliv matričních jevů na odlehlé hodnoty [30]. Obě studie tento závěr podporují.

Matriční jev je viditelný také v regresních grafech, kde je poloha bodů o stejné koncentraci, ale rozdílných ablačních energiích, rozmazána v rovině podílu intenzity a energie laseru, přestože by se podle předpokládané lineární závislosti měla shodovat. Směr relativního posunutí bodů v závislosti na ablační energii a relativní intenzitě vůči ostatním vzorkům je v souladu s tvrzením o nelinearitě způsobené matričním jevem.

Tabulka 4.1: Výsledky lineární regrese

(a) Cr

Energie (mJ)	RMSE	MAE	R ²
40	9.79	9.19	0.16
50	10.23	9.69	0.03
60	10.51	9.99	-0.02
70	10.8	10.29	-0.02
80	10.81	10.3	-0.02

(b) Mo

Energie (mJ)	RMSE	MAE	R ²
40	1.39	1.16	-0.31
50	1.31	1.09	-0.17
60	1.27	1.06	-0.10
70	1.27	1.05	-0.09
80	1.26	1.05	-0.08

(c) Ni

Energie (mJ)	RMSE	MAE	R ²
40	10.23	8.41	0.09
50	9.90	7.95	0.14
60	9.83	7.80	0.16
70	9.93	7.98	0.14
80	10.20	8.45	0.09

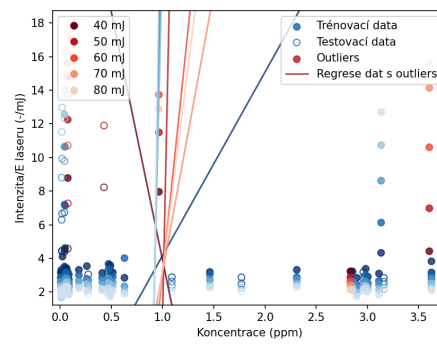
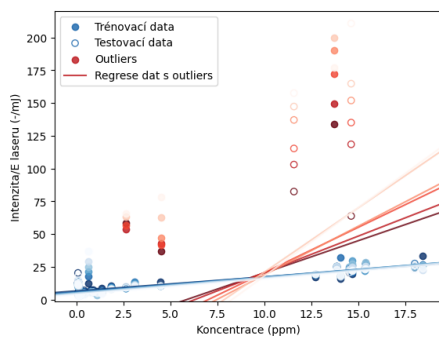
(d) Mn

Energie (mJ)	RMSE	MAE	R ²
40	0.48	0.42	0.41
50	0.59	0.50	0.12
60	0.65	0.55	-0.10
70	0.72	0.63	-0.34
80	0.70	0.61	-0.25

Obrázek 4.1: Výsledky lineární regrese

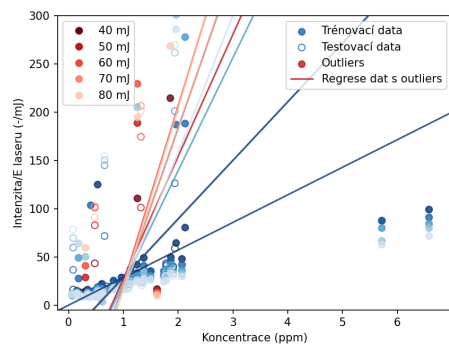
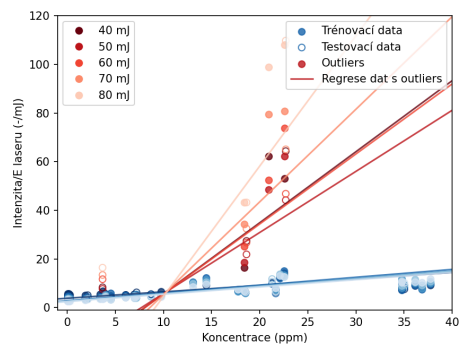
(b) Mo

(a) Cr



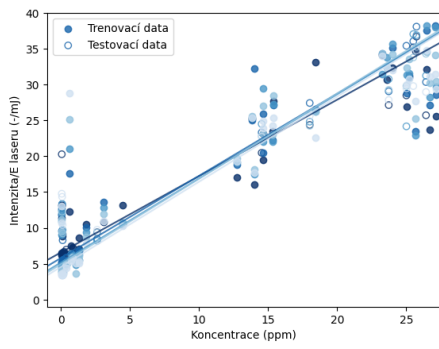
(c) Ni

(d) Mn

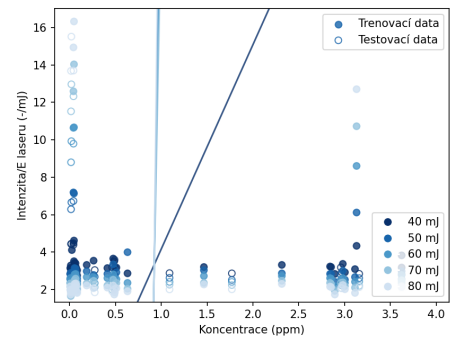


Obrázek 4.2: Lineární regrese po odstranění odlehlých hodnot

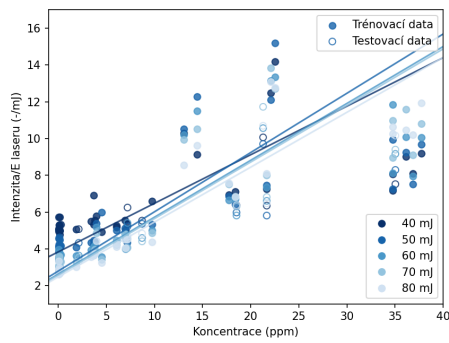
(a) Cr



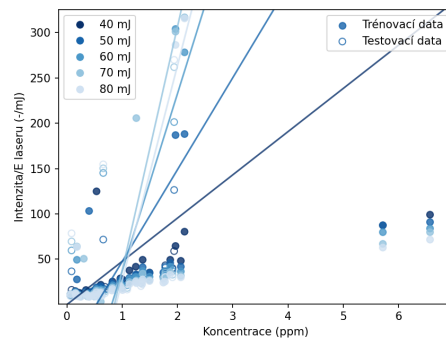
(b) Mo



(c) Ni



(d) Mn



Tabulka 4.2: Výsledky lineární regrese po odstranění odlehlých hodnot

(a) Cr				(b) Mo			
Energie (mJ)	RMSE	MAE	R ²	Energie (mJ)	RMSE	MAE	R ²
40	5.03	3.48	0.81	40	1.14	0.98	0.02
50	3.02	2.25	0.93	50	1.16	1.01	-0.01
60	4.00	3.01	0.88	60	1.19	1.05	-0.07
70	2.93	2.29	0.93	70	1.24	1.10	-0.17
80	4.41	3.39	0.85	80	1.22	1.08	-0.13

(c) Ni				(d) Mn			
Energie (mJ)	RMSE	MAE	R ²	Energie (mJ)	RMSE	MAE	R ²
40	9.01	6.67	0.37	40	0.59	0.17	0.17
50	8.21	5.86	0.48	50	0.71	0.59	-0.18
60	6.51	5.09	0.67	60	0.68	0.59	-0.11
70	6.33	4.98	0.69	70	0.75	0.63	-0.33
80	5.02	3.67	0.80	80	1.19	0.93	0.09

4.2 Stavba a optimalizace ANN

4.2.1 Optimalizace algoritmem hyperband

Provedl jsem optimalizaci hyperparametrů neuronové sítě metodou hyperband. Počty neuronů každé vrstvy jsem nechal generovat v rozmezí 5 až 200 s krokem 5. Learning rate jsem nechal generovat v rozmezí 10^{-1} až 10^{-5} bez kroku, ale s logaritmickým pravděpodobnostním rozložením. Každý model jsem nechal běžet po dobu maximálně 50 epoch. Vygenerované architektury jsem následně otestoval na celém trénovacím datasetu metodou 6-fold cross validace. Výsledky měření jsou uvedeny v Tabulce 4.3. Learning rate je zde zaokrouhlený na 1 platnou cifru. Architekturám se až na výjimky dařilo výrazně hůře než na datasetu, pro který byly selektovány. Příčinou byl overfitting a underfitting architektur na některých cross-validačních datasetech, což vedlo k velmi slabému celkovému RMSE. Přestože tento výsledek je v souladu s článkem [20], na špatném výkonu se pravděpodobně také podílela absence regularizačních prvků [31] a míra komplexity trénovacích dat v poměru k počtu vzorků. Článek [13] doporučuje v případě komplexní matrice použít variantu leave-one-out cross-validace, která je ovšem mnohonásobně výpočetně náročnější. Špatný výkon sítí nepředstavoval problém, neboť sítě stejně musely být naučně upraveny a obohaceny o regularizační prvky. Experiment mi poskytl dobrý základ pro další optimalizace a jednoznačně byl časově výhodnější

oproti případu, ve kterém bych musel od začátku postupovat manuálně.

Tabulka 4.3: Top 10 architektur generovaných metodou hyperband

Vrstva 1	Vrstva 2	LR	Val. Loss	CV RMSE	CV MAE
185	105	0.0006	0.20	3.11	1.77
80	130	0.0008	0.18	0.74	0.59
160	100	0.001	0.21	1.48	1.14
95	125	0.0009	0.16	20.58	8.75
100	105	0.0007	0.22	0.56	0.46
165	165	0.0009	0.19	1.54	1.07
105	125	0.0006	0.23	5.35	2.95
50	130	0.0008	0.20	1.09	0.97
20	125	0.001	0.18	4.53	2.50
75	60	0.001	0.17	10.87	4.84

4.2.2 Manuální optimalizace

Na základě nejvýkonnějších modelů vygenerovaných algoritmem hyperband jsem stanovil počet neuronů v obou hlubokých vrstvách jako 150. Dovolil jsem si zvolit vyšší hodnotu, aby model dokázal zachytit vztahy většího datasetu s komplexnější maticí, než jaký měl k dispozici algoritmus. Za účelem regularizace modelu jsem pod každou hlubokou vrstvu přidal dropout vrstvu. Z dropout vrstev jsem testoval kombinace hodnot 0, 0.1, 0.2, 0.3, 0.5. Pro každou kombinaci jsem vyzkoušel learning rate v rozmezí 10^{-3} , $3 \cdot 10^{-3}$, 10^{-4} , $3 \cdot 10^{-4}$, 10^{-5} . Experimentoval jsem také s batch size, kde jsem se každých 3 až 5 cyklů pokusil najít optimální hodnotu mezi mocninami čísla 2 od 1 do 16. Po gridu jsem postupoval po krocích a snažil se dojít k nejlepšímu lokálnímu řešení. Tím se ukázala být architektura zobrazená v Tabulce 4.4. Takovéto zvýšení komplexity však vedlo k prodloužení doby konvergence validační ztráty na 100 epoch. Velmi výrazným faktorem za výkonem sítě v průběhu trénování byl batch size 1. Možným důvodem je, že zvýšená stochastická způsobená nízkou batch size přidává tréninku jistý faktor regularizace. [32]

Pro tuto architekturu jsem následně natrénoval pět neuronových sítí pro každý prvek a každou energii ablačního laseru. Výsledky predikce sítí na respektivních testovacích datech jsou uvedeny v Tabulce 4.5).

V datech lze obecně pozorovat, že s rostoucí energií laseru se snižují chybové metriky testovaných neuronových sítí. Výjimkou je chrom (Cr), který

Tabulka 4.4: Architektura finální manuálně optimalizované ANN

Layer	Units/Rate	Parameters
Input	1957	1957
Dense 1	150	293700
Dropout 1	0.5	0
Dense 2	150	22650
Dropout 2	0.5	0
Output	150	151
Learning rate:	0.00003	
Batch size:	1	

vykazuje nárůst chybovosti při nejvyšší ablační energii. Fakt, že se toto chování vyskytuje opakovaně, a i u fine-tuned neuronové sítě, poukazuje na to, že se pravděpodobně nejedná o anomálii při trénování. Výchylka je tedy nejspíše důsledkem nasycení spektra nebo jiné složité vysokoenergetické interakce.

Pozitivní korelace výkonu s ablační energií je pravděpodobně následkem zvýšení poměru signálu k šumu. O tomto závěru také svědčí fakt, že nejvyšší zlepšení predikčních schopností neuronových sítí bylo pozorováno u prvků o nejnižší intenzitě silných spektrálních čar. Stejný efekt byl pozorován v následující literatuře: [33], [8]

Následně jsem zkoumal vliv změny ablační energie na natrénované neuronové sítě. K tomuto účelu jsem použil modely trénované na ablační energii 50 mJ a kalibroval s nimi spektra naměřená na ostatních energiích. Výsledky jsou uvedeny v Tabulce 4.6. Zatímco se výsledky v rámci jednotlivých prvků řídí velmi jasným a předvídatelným trendem, napříč prvky tomu tak není. Pro Cr a Ni klesá jak nepřesnost tak nejistota predikcí s rostoucí ablační energií, u Mo a Mn se tyto metriky zvyšují. Možným vysvětlením je zvýšení nelinearity spektrálních čar vlivem matričního jevu. Tato doměnka je v souladu s nálezy učiněnými v sekci lineární regrese, kde prvky Mn a Mo po odstranění odlehlých hodnot také vykazovaly nárůst RMSE s rostoucí ablační energií a prvek Ni vykazoval pokles.

4.3 Transfer learning

Cílem mojí práce bylo prozkoumat a adresovat vliv rozdílu v ablační energii na výkon kalibračních modelů u mnou vybraných metod. V tomto rámci jsem zkoumal MLP jako metodu kalibrace a transfer learning jako metodu adresace rozdílu v energii laseru. Jako zdrojový dataset pro transfer learning jsem používal spektra naměřená na ablační energii 50 mJ, jako fine tuning

Tabulka 4.5: Výkon architektury pro různé prvky a ablační energie

(a) Cr			(b) Mo		
Energie (mJ)	RMSE (ppm)	MAE (ppm)	Energie (mJ)	RMSE (ppm)	MAE (ppm)
40	1.86±0.25	1.40±0.14	40	0.81±0.10	0.52±0.04
50	1.21±0.18	0.79±0.11	50	0.76±0.09	0.44±0.03
60	1.64±0.24	1.10±0.13	60	0.71±0.08	0.44±0.03
70	1.35±0.20	0.85±0.09	70	0.48±0.06	0.31±0.03
80	1.16±0.17	0.76±0.08	80	0.37±0.05	0.27±0.02
90	1.51±0.22	0.94±0.10	90	0.37±0.05	0.27±0.02

(c) Mn			(d) Ni		
Energie (mJ)	RMSE (ppm)	MAE (ppm)	Energie (mJ)	RMSE (ppm)	MAE (ppm)
40	0.17±0.05	0.12±0.03	40	3.64±0.40	2.56±0.30
50	0.10±0.03	0.09±0.02	50	2.81±0.35	2.01±0.24
60	0.15±0.04	0.11±0.02	60	2.31±0.30	1.59±0.20
70	0.12±0.04	0.11±0.02	70	2.17±0.28	1.65±0.22
80	0.11±0.03	0.09±0.02	80	2.02±0.26	1.38±0.18
90	0.14±0.04	0.11±0.02	90	1.38±0.20	1.10±0.15

dataset všechna ostatní spektra. Pro každou z 5 sítí natrénovaných na energii 50 mJ jsem provedl fine-tuning jednou pro každou další ablační energii. Learning rate jsem pro každou síť použil stejný, jako při jejím učení, tedy $3 \cdot 10^{-4}$. Při použití batch size 1 dosahovaly sítě maxima zpravidla během jedné epochy a při validaci vykazovaly vysokou variaci. Přesnými hodnotami pro Cr se zabývám v druhé části experimentu. Na základě tohoto zjištění jsem experimentoval s rozdílnými batch sizes pro prvky Cr, Mn a ablační energie 40 mJ, 80 mJ. Nakonec jsem zvolil batch size 32, jelikož nesla nejlepší výsledky napříč testovanými prvky a energiemi laseru. U všech prvků vyjma Mo docházelo ke konvergenci velmi konzistentně mezi epochami 7 a 10, dále u nich nedocházelo k významnému přetrénování. Pro Mo docházelo ke konvergenci kolem 17. epochy. Neuronové sítě chromu, niklu a manganu jsem proto trénoval celkem 10 epoch, neuronové sítě molybdenu 20 epoch. Výsledky predikcí těmito neuronovými sítěmi uvádím v Tabulce 4.7.

Hodnotící kritéria jednotlivých měření jsem vnesl do Grafů 4.3, 4.4. Z grafů nelze určit souvislý trend. V případě chromu vykazují lepší výkon neu-

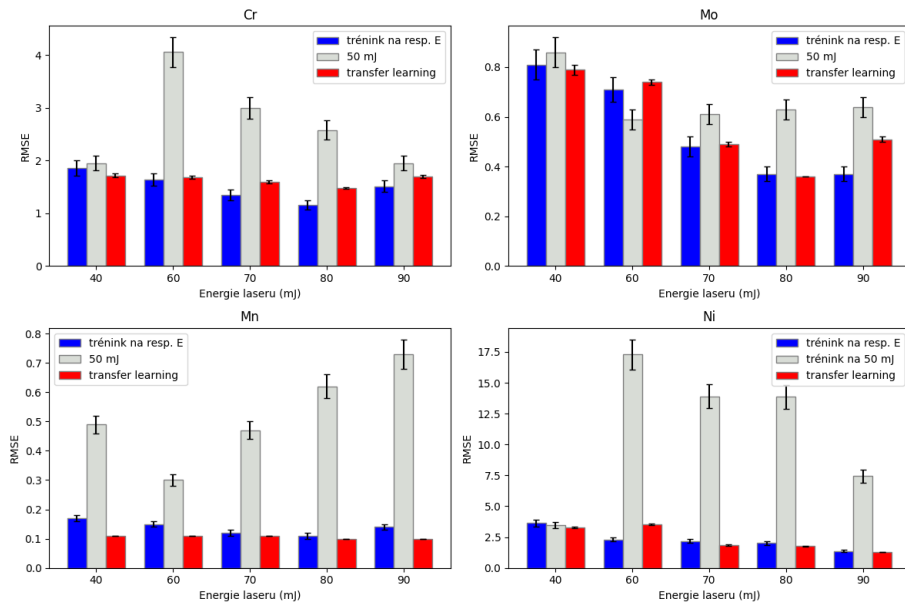
Tabulka 4.6: Výkon ANN trénovaných na energii 50 mJ na ostatních energiích

(a) Cr			(b) Mo		
Energie (mJ)	RMSE (ppm)	MAE (ppm)	Energie (mJ)	RMSE (ppm)	MAE (ppm)
40	1.95±0.14	1.68±0.12	40	0.86±0.06	0.51±0.04
60	4.06±0.28	3.05±0.21	60	0.59±0.04	0.41±0.03
70	3.00±0.21	2.43±0.17	70	0.61±0.04	0.41±0.04
80	2.58±0.18	2.17±0.15	80	0.63±0.04	0.42±0.03
90	1.95±0.14	1.68±0.12	90	0.64±0.04	0.42±0.03

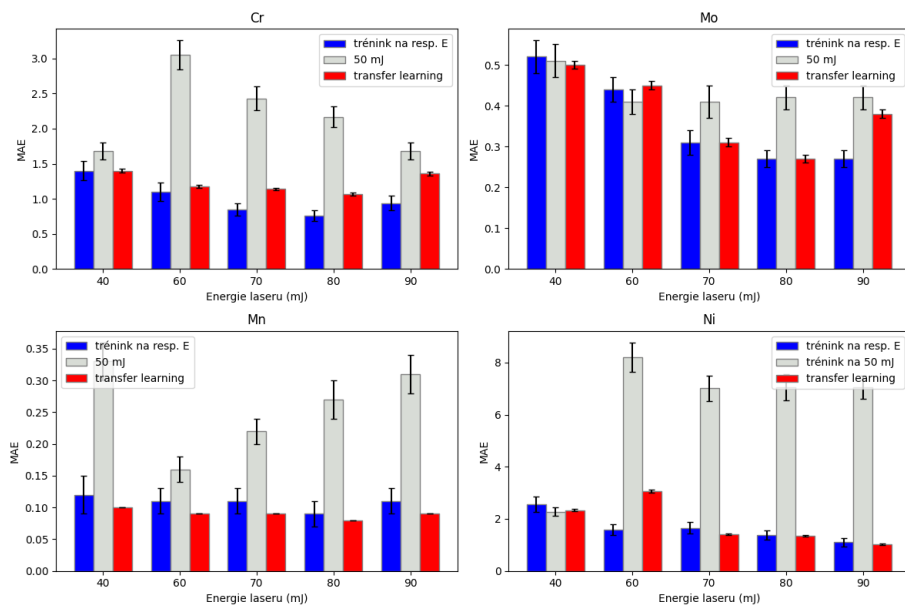
(c) Mn			(d) Ni		
Energie (mJ)	RMSE (ppm)	MAE (ppm)	Energie (mJ)	RMSE (ppm)	MAE (ppm)
40	0.49±0.03	0.33±0.03	40	3.46±0.24	2.27±0.16
60	0.30±0.02	0.16±0.02	60	17.30±1.21	8.20±0.57
70	0.47±0.03	0.22±0.02	70	13.90±0.97	7.01±0.49
80	0.62±0.04	0.27±0.03	80	13.86±0.97	7.05±0.49
90	0.73±0.05	0.31±0.03	90	7.44±0.52	7.05±0.45

ronové sítě trénované od základu na svých respektivních energiích. V případě manganu mají lepší výsledky fine-tuned neuronové sítě. Výsledky niklu a molybdenu jsou nejednoznačné a zpravidla spadají do svých vzájemných intervalů spolehlivosti. Fine-tuned neuronové sítě mají obecně značně nižší směrodatné odchylky. To může být následkem redukce stochasticity z důvodu navýšení batch size nebo dotrénování.

Kritéria



Obrázek 4.3: Porovnání výsledků RMSE



Obrázek 4.4: Porovnání výsledků MAE

Tabulka 4.7: Výkon fine-tuned neuronových sítí

(a) Cr

Energie (mJ)	RMSE (ppm)	MAE (ppm)
40	1.72 ± 0.03	1.40 ± 0.03
60	1.69 ± 0.03	1.18 ± 0.02
70	1.60 ± 0.03	1.14 ± 0.02
80	1.48 ± 0.02	1.07 ± 0.02
90	1.70 ± 0.03	1.36 ± 0.03

(b) Mo

Energie (mJ)	RMSE (ppm)	MAE (ppm)
40	0.79 ± 0.02	0.50 ± 0.01
60	0.74 ± 0.01	0.45 ± 0.01
70	0.49 ± 0.01	0.31 ± 0.01
80	0.36 ± 0.00	0.27 ± 0.01
90	0.51 ± 0.01	0.38 ± 0.01

(c) Mn

Energie (mJ)	RMSE (ppm)	MAE (ppm)
40	0.11 ± 0.00	0.10 ± 0.00
60	0.11 ± 0.00	0.09 ± 0.00
70	0.11 ± 0.00	0.09 ± 0.00
80	0.10 ± 0.00	0.08 ± 0.00
90	0.10 ± 0.00	0.09 ± 0.00

(d) Ni

Energie (mJ)	RMSE (ppm)	MAE (ppm)
40	3.27 ± 0.07	2.33 ± 0.05
60	3.55 ± 0.07	3.05 ± 0.06
70	1.85 ± 0.04	1.42 ± 0.03
80	1.77 ± 0.03	1.35 ± 0.03
90	1.30 ± 0.02	1.02 ± 0.02

Kapitola 5

Závěr

Cílem mojí práce bylo prozkoumat a adresovat vliv rozdílu v ablační energii na výkon kalibračních modelů u mnou vybraných metod. V tomto rámci jsem zkoumal MLP jako metodu kalibrace a transfer learning jako metodu korekce rozdílu v ablační energii.

V rámci práce jsem navrhnul architekturu neuronové sítě pro kalibraci spekter 4 různých prvků - chromu, niklu, molybdenu, manganu, a 6 různých ablačních energií - 40, 50, 60, 70, 80 a 90 mJ. Pro každý prvek a každou ablační energii jsem natrénoval 5 modelů s touto architekturou a nechal je provádět predikce na dříve neviděných datech. V návaznosti na to jsem použil modely natrénované na ablační energii 50 mJ k predikci koncentrace prvků ze spekter měřených na ostatních ablačních energiích. Rozdíly v měření jsem se následně pokusil kompenzovat aplikací transfer learningu.

Z experimentu vyplývá, že transfer learning může být validní metodou pro korekci nepřesností způsobených rozdílem ablačních energií. Prokázal jsem jeho schopnost dosáhnout srovnatelných výsledků s modely trénovanými od základů, ale za zlomek času a s výrazně nižší výpočetní náročností. V rámci práce se mi však nepodařilo prokázat konsistentní zlepšení výkonu neuronových sítí použitím transfer learningu v kontextu LIBS. Na vině je pravděpodobně nedostatečná optimalizace použitých neuronových sítí, nedostatečná komplexita experimentálního datasetu, případně jistá kombinace těchto faktorů.

Dále se mi podařilo demonstrovat obtížnou předvídatelnost vlivu změny ablačních energií na neuronové sítě, s jedinečnými reakcemi každého testovaného prvku na změnu ablačních energií.

Literatura

- [1] M. Sanz *et al.*, “Laser-induced breakdown spectroscopy: A tool for real-time, in situ, and remote sensing of elemental compositions,” *ChemInform*, vol. 50, 2019.
- [2] J. Anzano, S. López, and J. Laserna, “Laser-induced breakdown spectroscopy: Fundamentals, applications, and challenges,” *ISRN Spectroscopy*, 2013.
- [3] M. Markiewicz-Keszycka, R. Cama-Moncunill, M. P. Casado-Gavaldà, Y. Dixit, X. Cama-Moncunill, and C. Sullivan, “Laser-induced breakdown spectroscopy (libs) for food analysis: a review,” *Trends in Food Science Technology*, vol. 65, pp. 80–93, 2017.
- [4] S. M. Clegg, R. C. Wiens, R. B. Anderson, O. Forni, M. D. Dyar, J. Cook, R. Newell, S. K. Sharma, P. Sobron, and S. A. Mertzman, “Miniaturized laser heterodyne radiometer for measurements of CO₂ from space: Validation on atmospheric samples,” *Applied Optics*, vol. 48, no. 4, pp. B165–B171, 2009.
- [5] D. W. Hahn and N. Omenetto, “Laser-induced breakdown spectroscopy (libs), part i: Review of basic diagnostics and plasma-particle interactions: Still-challenging issues within the analytical plasma community,” *Applied Spectroscopy*, vol. 64, no. 12, pp. 335A–366A, 2010.
- [6] G. Galbács, “The principles of and reasons for using laser-induced breakdown spectrometry: an overview,” *Microchemical Journal*, vol. 124, pp. 31–39, 2015.
- [7] S. Maurice, R. Wiens, M. Saccoccio *et al.*, “The chemcam instrument suite on the Mars Science Laboratory (MSL) rover: Science objectives and mast unit description,” *Space Science Reviews*, vol. 170, pp. 95–166, 2012.

- [8] I. B. Gornushkin, C. L. Stevenson, G. Galbács, B. W. Smith, and J. D. Winefordner, “Multivariate calibration of laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and artificial neural network analysis methods,” *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 55, no. 10, pp. 1603–1630, 2000.
- [9] O. A. Al-Najjar, Y. S. Wudil, U. F. Ahmad, O. S. B. Al-Amoudi, M. A. Al-Osta, and M. A. Gondal, “Applications of laser induced breakdown spectroscopy in geotechnical engineering: a critical review of recent developments, perspectives and challenges,” *Applied Spectroscopy Reviews*, vol. 0, no. 0, pp. 1–37, 2022. [Online]. Available: <https://doi.org/10.1080/05704928.2022.2136192>
- [10] A. W. Miziolek, V. Palleschi, and I. Schechter, “Laser-induced breakdown spectroscopy (libs), part ii: Review of instrumental and methodological approaches to material analysis and applications to different fields,” *Applied Spectroscopy Reviews*, vol. 41, no. 4, pp. 331–366, 2006.
- [11] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013.
- [12] M. Sokolova and G. Lapalme, *Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation*. Springer Berlin Heidelberg, 2006.
- [13] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, vol. 2. Morgan Kaufmann Publishers Inc., 1995, pp. 1137–1143.
- [14] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006. [Online]. Available: <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>
- [15] G. C. Cawley and N. L. Talbot, “On over-fitting in model selection and subsequent selection bias in performance evaluation,” *Journal of Machine Learning Research*, vol. 11, no. Jul, pp. 2079–2107, 2010. [Online]. Available: <https://jmlr.org/papers/v11/cawley10a.html>
- [16] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, “Efficient backprop,” *Neural networks: Tricks of the trade*, vol. 7700, pp. 9–48, 2012.

- [17] D. S. Lolla, S. Roy, and S. K. Saha, “Hyperparameter optimization in machine learning models,” in *Handbook of Research on Machine and Deep Learning Applications in Science and Engineering*, S. Patnaik, P. K. Pattnaik, R. C. Poonia, and M. Jena, Eds. IGI Global, 2020, pp. 1–29.
- [18] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *Journal of Machine Learning Research*, vol. 13, no. Feb, pp. 281–305, 2012.
- [19] L. Li, K. G. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, “Efficient hyperparameter optimization and infinitely many armed bandits,” *CoRR*, vol. abs/1603.06560, 2016. [Online]. Available: <http://arxiv.org/abs/1603.06560>
- [20] K. Kandasamy, J. Schneider, and B. Póczos, “Parallelised bayesian optimisation via thompson sampling,” *International Conference on Artificial Intelligence and Statistics*, pp. 133–142, 2018.
- [21] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [22] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” in *Advances in neural information processing systems*. MIT Press, 2014, pp. 3320–3328.
- [23] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” in *International conference on artificial neural networks*. Springer, 2018, pp. 270–279.
- [24] E. Kepes, “test_dataset_raw,” 7 2022. [Online]. Available: https://figshare.com/articles/dataset/test_dataset_RAW/20297136
- [25] “LIBS 2022 Regression Contest,” 2022, <https://libs2022.com/index.php/regression-contest/>.
- [26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: <http://jmlr.org/papers/v15/srivastava14a.html>

- [27] AtomTrace, “Atomtrace spectral database,” <https://www.atomtrace.com/elements-database/>, 2023.
- [28] I. Traparić and M. Ivković, “Determination of austenitic steel alloys composition using laser-induced breakdown spectroscopy (libs) and machine learning algorithms,” *Eur. Phys. J. D*, vol. 77, p. 30, 2023. [Online]. Available: <https://doi.org/10.1140/epjd/s10053-023-00608-6>
- [29] D. A. Cremers and L. J. Radziemski, *Handbook of Laser-Induced Breakdown Spectroscopy*. John Wiley & Sons, Ltd, 2006. [Online]. Available: <https://www.wiley.com/en-us/Handbook+of+Laser+Induced+Breakdown+Spectroscopy%2C+2nd+Edition-p-9781118567371>
- [30] J. A. Aguilera and C. Aragón, “Characterization of laser induced plasmas by optical emission spectroscopy: A review of experiments and methods,” *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 59, no. 9, pp. 1865–1875, 2004.
- [31] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 2009. [Online]. Available: <https://web.stanford.edu/~hastie/ElemStatLearn/>
- [32] A. C. Wilson and T. R. Martinez, “The general inefficiency of batch training for gradient descent learning,” *Neural Networks*, vol. 16, no. 1, pp. 142–154, 2003.
- [33] J. L. Gottfried, “Influence of er: Yag and nd: Yag wavelengths on laser-induced breakdown spectroscopy measurements under air and helium atmosphere,” *Applied optics*, vol. 46, no. 19, pp. 3966–3977, 2007.