



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

FORENZNÍ VERIFIKACE AUTORA RUČNĚ PSANÉHO ZNAKU "K" S POMOCÍ UMĚLÉ INTELIGENCE

FORENSIC VERIFICATION OF AUTHOR OF HANDWRITTEN CHARACTER "K" USING ARTIFICIAL
INTELLIGENCE

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Eliška Bachorecová

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Ing. Radim Burget, Ph.D.

BRNO 2025



Bakalářská práce

bakalářský studijní program **Informační bezpečnost**

Ústav telekomunikací

Studentka: Eliška Bachorecová

ID: 247514

Ročník: 3

Akademický rok: 2024/25

NÁZEV TÉMATU:

Forenzní verifikace autora ručně psaného znaku "k" s pomocí umělé inteligence

POKYNY PRO VYPRACOVÁNÍ:

Ekonomická i násilná kriminalita je často spojena s dokazováním autorství ručně psaných dokumentů. S příchodem digitálních technologií se však množství ručně psaného textu snižuje. Cílem práce je připravit trénovací a testovací databázi znaku "k" a navrhnout dvě či více architektur neuronové sítě pro verifikaci autora. Prvním krokem bude sběr dostatečného množství vzorků ručně psaného znaku "k" od různých autorů. Následně navrhnete a implementujete různé architektury neuronových sítí, které budou schopny verifikovat autora na základě analýzy těchto vzorků. Dosažené výsledky srovnajte pomocí vhodných metrik, jako jsou přesnost, F1 skóre nebo AUC. Závěrem porovnáte výsledky vašeho výzkumu s aktuálním stavem vědy a techniky ve světě, čímž zhodnotíte jejich efektivitu a přínosnost.

DOPORUČENÁ LITERATURA:

Podle pokynů vedoucího práce.

Termín zadání: 10.2.2025

Termín odevzdání: 3.6.2025

Vedoucí práce: doc. Ing. Radim Burget, Ph.D.

prof. Ing. Jan Hajný, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Bakalářská práce se zabývá návrhem a implementací tří architektur neuronových sítí, jakožto podpůrných nástrojů v oblasti expertizy ručního písma. Účelem modelů je na základě komparace variací referenčních a sporných znaků „k“ stanovit, zda jejich autorem je tentýž člověk, či nikoliv. Datové množiny vytvořené odborným konzultantem obsahují také vzorky písmene „a“, které jsou zahrnuty do zpracování. Teoretická část práce se věnuje analýze aktuálního stavu expertizy ručního písma a identifikaci klíčových problémů, se kterými se tato disciplína v současnosti potýká. Následuje srovnání konvenčních a moderních přístupů, dle kterého jsou vybrány relevantní technologie. Pro účely práce je navržena konvoluční neuronová síť a siamské síť. Praktická část práce obsahuje představení datových množin a následný popis implementovaných architektur. Hodnoty evaluačních metrik jsou pro jednotlivé modely vyneseny do odpovídajících tabulek a dosažené výsledky jsou patřičně vizualizovány. Na závěr je tato práce porovnána s ostatními studii, jež reflektují aktuální stav vědy.

Na testovacích datových množinách tvořených 180 vzorky písmene „k“ a 175 vzorky písmene „a“ od 45 autorů je dosaženo *přesnosti* 93,4 % a 95,7 %, a to s využitím zóny nejistoty. Ze zhotoveného dotazníku, jehož cílem bylo ověřit praktickou využitelnost navržených metod, vyplývá, že model je schopen verifikovat autora s *vyšší přesností* než laická veřejnost, a to v mnohonásobně kratším čase. Na rozdíl od majoritní části studií zde dochází ke klasifikaci na úrovni grafému s využitím datových množin obsahujících autentické vzorky, čímž je cíleno na praktickou využitelnost.

KLÍČOVÁ SLOVA

Expertiza ručního písma, hluboké učení, konvoluční neuronová síť, siamská síť, verifikace autora

ABSTRACT

The bachelor thesis deals with design and implementation of three neural networks architectures as a support tool in the field of forensic handwriting examination. The purpose of the models is to determine, based on a comparison of variations of the reference and disputed characters “k”, whether their author is the same person or not. The datasets created by a consultant also contain samples of the letter “a”, which are included in the processing. The theoretical part is devoted to the analysis of forensic handwriting examination and identification of key challenges the discipline currently faces. Next, relevant technologies are selected based on a comparison of conventional and modern approaches. For the purpose of this thesis a convolutional neural network and siamese networks are proposed. The practical part introduces datasets and provides description of implemented architectures. The values of the evaluation metrics for the individual models are presented in the corresponding tables and the achieved results are appropriately visualized. Finally, this thesis is compared with other studies that reflect the current state of the art.

On test datasets consisting of 180 samples of the letter “k” and 175 samples of the letter “a” from 45 authors, *accuracies* of 93.4 % and 95.7 % are achieved, using the uncertainty zone. The results of conducted questionnaire, which aimed at verifying practical applicability of the proposed methods, show that the model is able to verify the author with higher *accuracy* than lay public and in significantly less time. Unlike the majority of studies, classification is performed here at grapheme level using datasets containing authentic samples, thereby aiming for practical applicability.

KEYWORDS

Forensic handwriting examination, deep learning, convolutional neural network, siamese network, author verification

BACHORECOVÁ, Eliška. *Forenzní verifikace autora ručně psaného znaku "k" s pomocí umělé inteligence*. Bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2025. Vedoucí práce: doc. Ing. Radim Burget, Ph.D.

Prohlášení autora o původnosti díla

Jméno a příjmení autora:	Eliška Bachorecová
VUT ID autora:	247514
Typ práce:	Bakalářská práce
Akademický rok:	2024/25
Téma závěrečné práce:	Forenzní verifikace autora ručně psaného znaku "k" s pomocí umělé inteligence

Prohlašuji, že svou závěrečnou práci jsem vypracovala samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autorky*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucímu bakalářské práce panu doc. Ing. Radimu Burgetovi, Ph.D. za odborné vedení, konzultace a podnětné návrhy k práci. Dále bych chtěla poděkovat své rodině a přátelům za jejich podporu, trpělivost a pochopení.

Obsah

Úvod	12
1 Teoretická část bakalářské práce	14
1.1 Expertiza ručního písma v kriminalistice	14
1.2 Postup verifikace autora ručně psaného textu	16
1.2.1 Sběr dat	16
1.2.2 Digitalizace snímků	16
1.2.3 Předzpracování dat	17
1.2.4 Extrahování příznaků a klasifikace	18
1.2.5 Zhodnocení algoritmů	18
1.3 Konvenční metody	20
1.4 Moderní metody	21
1.4.1 Konvoluční neuronové sítě	22
1.4.2 Siamské sítě	25
1.4.3 Transformery	26
1.5 Přehled známých architektur	28
1.5.1 Architektura VGG	28
1.5.2 Architektura ResNet	29
1.5.3 Architektura EfficientNet	29
1.6 Teplotní mapy	30
2 Praktická část bakalářské práce	31
2.1 Datová množina	31
2.2 Navržené metody určení shody	32
2.2.1 Model konvoluční neuronové sítě	33
2.2.2 Model siamské sítě s autoenkodérem	35
2.2.3 Model konvoluční siamské sítě	37
2.3 Dosažené výsledky	39
2.3.1 Evaluace modelu konvoluční neuronové sítě	40
2.3.2 Evaluace modelu siamské sítě s autoenkodérem	43
2.3.3 Evaluace modelu konvoluční siamské sítě	46
2.3.4 Výstupy teplotních map	50
2.3.5 Ověření praktické využitelnosti práce	52
2.4 Výsledky práce v kontextu současného stavu vědy	53
Závěr	55
Literatura	57

Seznam symbolů a zkratk	64
A Obsah elektronické přílohy	66

Seznam obrázků

1.1	Demonstrace variability v rámci pisatele	15
1.2	Obecné schéma CNN	22
1.3	Operace konvoluce	23
1.4	Obecné schéma siamské sítě	25
1.5	Architektura ViT	27
2.1	Obecné schéma experimentu	31
2.2	Předzpracování dat pro CNN	33
2.3	Architektura CNN	34
2.4	Předzpracování dat pro siamskou síť s autoenkodérem	35
2.5	Architektura siamské sítě s autoenkodérem	36
2.6	Autoenkodér – součást siamské sítě	37
2.7	Předzpracování dat pro konvoluční siamskou síť	37
2.8	Architektura konvoluční siamské sítě	39
2.9	Průběhy jednotlivých trénování CNN	40
2.10	Matice záměn CNN pro znak „k“	41
2.11	Matice záměn CNN pro znak „a“	41
2.12	Matice záměn CNN pro kombinaci „k“ a „a“	41
2.13	ROC křivky CNN pro znak „k“	42
2.14	ROC křivky CNN pro znak „a“	42
2.15	ROC křivky CNN pro kombinaci „k“ a „a“	42
2.16	Průběhy jednotlivých trénování siamské sítě s autoenkodérem	43
2.17	Matice záměn siamské sítě s autoenkodérem pro znak „k“	44
2.18	Matice záměn siamské sítě s autoenkodérem pro znak „a“	44
2.19	Matice záměn siamské sítě s autoenkodérem pro kombinaci „k“ a „a“	44
2.20	ROC křivky siamské sítě s autoenkodérem pro znak „k“	45
2.21	ROC křivky siamské sítě s autoenkodérem pro znak „a“	45
2.22	ROC křivky siamské sítě s autoenkodérem pro kombinaci „k“ a „a“	45
2.23	Průběhy jednotlivých trénování konvoluční siamské sítě	47
2.24	Matice záměn konvoluční siamské sítě pro znak „k“	48
2.25	Matice záměn konvoluční siamské sítě pro znak „a“	48
2.26	Matice záměn konvoluční siamské sítě pro kombinaci „k“ a „a“	48
2.27	ROC křivky konvoluční siamské sítě pro znak „k“	49
2.28	ROC křivky konvoluční siamské sítě pro znak „a“	49
2.29	ROC křivky konvoluční siamské sítě pro kombinaci „k“ a „a“	49
2.30	Teplotní mapa CNN	50
2.31	Teplotní mapa siamské sítě s autoenkodérem	51
2.32	Teplotní mapa konvoluční siamské sítě	51

Seznam tabulek

2.1	Datové množiny pro znak „k“	32
2.2	Datové množiny pro znak „a“	32
2.3	Metriky CNN pro písmena „k“ a „a“	40
2.4	Metriky siamské sítě s autoenkodérem pro písmena „k“ a „a“	43
2.5	Metriky konvoluční siamské sítě pro písmena „k“ a „a“ – průměr . .	47
2.6	Metriky konvoluční siamské sítě pro písmena „k“ a „a“ – majorita . .	47
2.7	Výsledky dotazníku srovnávajícího výstupy modelu, laiků a písmonalce	53

Úvod

Ručně psané písmo reprezentuje jedinečný projev lidské individuality, který nese nejen samotné sdělení, ale také množství doplňujících informací o autorovi. Při tvorbě rukopisu dochází k integraci neurologických, fyziologických, sensorických a psychologických faktorů [1]. Komplexnost tohoto procesu obohacuje samotnou informační hodnotu textu o specifické příznaky, jež jsou cenné pro odborníky z mnoha oborů. Analýza ručně psaného projevu může pomoci například při odhalení neurodegenerativních onemocnění [2]. Zásadní roli však sehrává v kriminalistice, kde se využívá k detekci falzifikátů, identifikaci či verifikaci autora a přispívá k sestavení psychologického profilu pachatele [3]. Informace, které poskytuje, mohou být důležitým důkazním materiálem v soudních řízeních [4]. Konkrétně se jedná o případy týkající se například rozvodového a opatrovnického řízení, majetkových sporů nebo pojistných podvodů [5].

Jednou z největších současných výzev v oblasti expertizy ručního písma je výrazný úbytek ručně psaných textů, což klade zvýšené nároky na kvalitu těchto expertiz. Hlavní příčinou úbytku psaného projevu je především digitalizace společnosti. Technologický pokrok však nepředstavuje pouze překážku, ale naopak může být využit jako efektivní nástroj k řešení současných problémů této disciplíny, jako je subjektivita a absence kvantitativních technik [6].

Tato práce vznikla v rámci kooperace s Kriminologickým ústavem v Praze. Jejím cílem je návrh, implementace a následná evaluace automatizovaných technik na bázi umělé inteligence určených ke komparaci variací referenčního a sporného písmene „k“. Výstupem tedy bude verifikace autora na úrovni grafému. Účelem práce není nahrazení lidského faktoru v oblasti expertizy ručního písma, ale tvorba podpůrného nástroje pro odborníky, kteří tyto expertizy provádí.

Hlavním přínosem této práce je návrh inovativních metod a prokázání, že ke stanovení, či vyloučení shody mezi referenčním a sporným rukopisem postačují pouze variace znaku „k“. Navržené metody dosahují, s využitím zóny nejistoty, *přesnosti* 93,4 %. Tato hodnota byla stanovena na vyvážené testovací datové množině obsahující 180 vzorků od 45 autorů. Práce je rovněž rozšířena i o výsledky dosažené na vzorcích písmene „a“, které odborný konzultant poskytl spolu s datovou množinou znaku „k“.

Stěžejním je využití autentické datové množiny obsahující grafém psaný jak tiskacím, tak psacím písmem. Tyto podmínky více korespondují s reálnými důkazními materiály, jež bývají k dispozici. Majoritní část studií zabývajících se touto problematikou je orientována spíše na zpracování celých vět, popřípadě slov získaných z velkých datových množin. Jejich dalším nedostatkem bývá využití datové množiny vytvořené za kontrolovaných podmínek.

Bakalářská práce je rozdělena na teoretickou a praktickou část. Teoretická část se zaměřuje na představení oboru expertizy ručního písma a analýzu aktuálních nedostatků, s nimiž se tato disciplína potýká. Následuje popis konvenčních a moderních přístupů v oblasti zpracování offline písma. Praktická část se věnuje návrhu a implementaci architektur dle zadání. V závěru jsou získané poznatky shrnuty, zhodnoceny a vizualizovány vhodnými metodami. Současně byl také zhotoven dotazník za účelem prokázání potenciální praktické využitelnosti navržených přístupů. V poslední kapitole je práce srovnána s ostatními studii, jež odpovídají současnému stavu vědy.

1 Teoretická část bakalářské práce

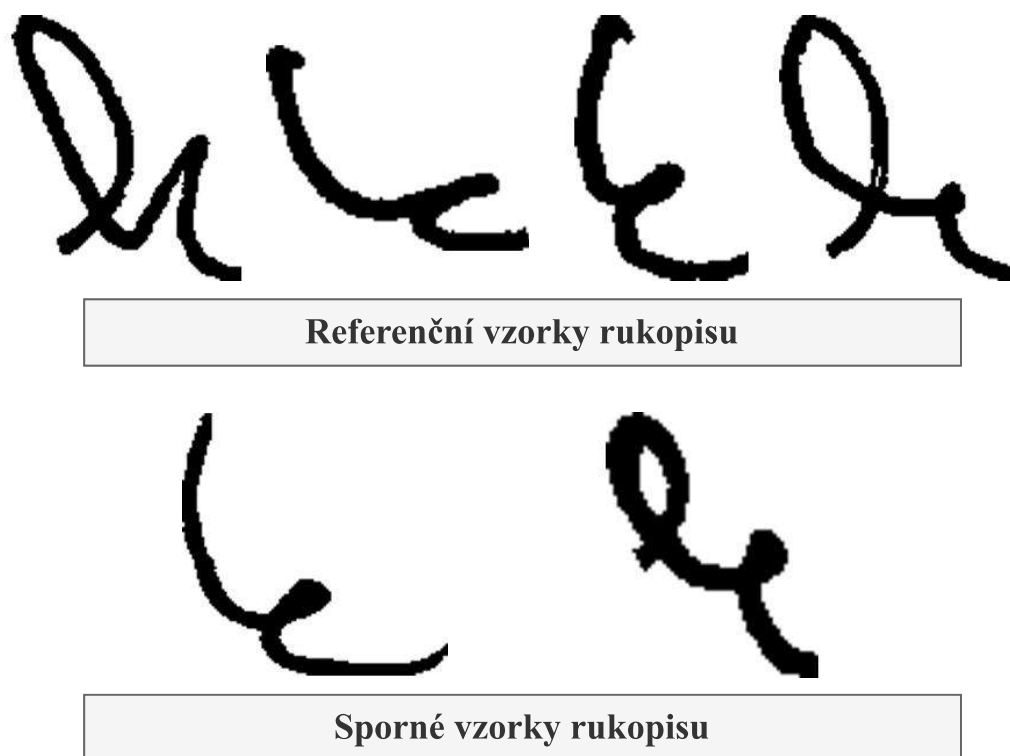
Tato část bakalářské práce analyzuje teoretická východiska v oblasti expertizy ručního písma. Dále jsou zde rozebrány principy a technologie uplatňované jak v konvenčních, tak moderních přístupech.

1.1 Expertiza ručního písma v kriminalistice

Expertiza ručního písma představuje metodu kriminalistické identifikace, jejímž cílem je ztotožnění pisatele. Podstatou celého procesu je srovnání referenčního a sporného materiálu [7], přičemž tato práce se výhradně zaměřuje na zjištění, zda autorem referenčního a sporného vzorku je tentýž člověk. Odborník detailně analyzuje specifické příznaky jednotlivých znaků, mezi které patří například stínování, výška, sklon, pravidelnost. Na základě jejich kombinace určuje, zda existují dostatečné podobnosti k potvrzení či vyvrácení autorství. Výchozí premisou je skutečnost, že žádné dvě osoby, včetně dvojčat, nepíší stejně [8]. Dle [4] je dominantní metodou expertizy ručního písma přístup známý jako ACE-V, jenž obsahuje fázi analýzy, srovnání, evaluace a následného ověření znalcem.

Jedná se o poměrně náročnou disciplínu. Jednou z příčin je vysoká variabilita grafému, viz obrázek 1.1. Člověk není schopen napsat jakýkoliv text dvakrát po sobě identickým způsobem [9]. Písmo se může lišit i v rámci slov. Z těchto důvodů se odborníci nikdy nemůžou rozhodovat na základě stoprocentní shody mezi referenčním a sporným vzorkem. Proměnlivost rukopisu je ovlivněna mnoha faktory, mezi které spadá vzdělání, národnost, psychické rozpoložení jedince, vyspělost. Současně obsahuje natolik specifické prvky, že pouze v extrémních případech je text neidentifikovatelný, což může být způsobeno volbou psacího nástroje nebo podkladového materiálu [10].

Za hlavní nedostatek expertizy ručního písma bývá považována její subjektivní povaha [11]. Veškeré závěry jsou stanoveny na základě kognitivních a percepčních schopností, což klade vysoké požadavky na zkušenost a odbornost znalců. Dále pak chybí kvantitativní srovnávací techniky [12]. Nevýhodou je také absence jednotné klasifikační stupnice, přičemž rozhodnutí expertů odpovídají kategorickým závěrům, které lze obtížně statisticky zpracovat [4]. Se záměrem minimalizovat či eliminovat tyto nedostatky, se vyvíjí nové technologie využívající benefity výpočetní techniky. Tyto inovativní přístupy vykazují potenciál nejen z hlediska kvantitativního zpracování dat, ale i v poskytování objektivních metrik umožňujících reprodukovatelnost výsledků. Dalším přínosem je výrazná časová úspora, jelikož manuální tvorba expertiz představuje zdlouhavý proces [13].



Obr. 1.1: Demonstrace variability v rámci pisatele

Již prvotní nástroje, jako byly CEDAR-FOX nebo FISH, vykazují jistý přínos pro tento obor [6]. Například zdroj [14] srovnává možné přístupy integrace technologií v oblasti ověřování offline podpisů, což je jedna z oblastí expertizy ručního písma. Mezi uvedené příklady patří metoda podpůrných vektorů (Support Vector Machines, SVM) nebo algoritmus k -nejbližších sousedů (k -Nearest Neighbors, k -NN) v kombinaci s lokálním binárním vzorem (Local Binary Pattern, LBP). Vysoký potenciál vykazují především přístupy založené na hlubokém učení, konkrétně konvoluční neuronové sítě. Metody byly testovány například na CEDAR [15] či MCYT [16] databázích, jež obsahují jak autentické, tak padělané podpisy. V rámci této práce budou dále rozebrány konvenční metody a podrobně analyzovány moderní přístupy vycházející z principů hlubokého učení.

1.2 Postup verifikace autora ručně psaného textu

Tato kapitola poskytuje obecný rámec pro verifikaci pisatele referenčního a sporného materiálu prostřednictvím výpočetních technologií. Postup vychází ze zdrojů [8, 13, 17, 18, 19]. Obsahuje celkem šest kroků, které je vždy nutno optimalizovat dle konkrétního případu. Ačkoliv se tato práce zaměřuje primárně na verifikaci autora, je zde čerpáno i ze zdrojů zabývajících se jeho identifikací. Hlavním důvodem je skutečnost, že obě tyto disciplíny si jsou velmi blízké a vychází ze stejného principu. Cílem verifikace je stanovit, zda dva rukopisy pochází od totožného autora, či nikoliv. Identifikace směřuje ke stanovení konkrétního pisatele na základě srovnávání vzorku psaného projevu s rukopisy z množiny potenciálních autorů. Obě disciplíny jsou tedy založeny na komparaci referenčního a sporného grafému, avšak podstata problematiky verifikace je binární, kdežto u identifikace je vícetřídní [17].

1.2.1 Sběr dat

Cílem této fáze je vytvoření adekvátních datových množin, a to trénovací, validační, testovací. Data mohou být získána fyzickým sběrem vzorků, popřípadě lze čerpat, z již zmíněných databází jako CEDAR či MCYT. Při volbě datové množiny je však nezbytné brát v potaz její specifické charakteristiky ovlivněné proměnlivostí didaktiky psaní v různých zemích. Velikost a kvalita těchto množin může výrazně ovlivnit následující kroky tohoto procesu. Malá trénovací datová množina či šum převažující nad reprezentativními daty mohou být jednou z příčin přetrénování neuronových sítí, což je stav, kdy model ztrácí schopnost adekvátně reagovat na nová data [20]. Dalším důležitým aspektem je, zda rukopisy vznikly za kontrolovaných podmínek, čímž může být pozmeněna přirozená variabilita pisatele. Takto opatřená datová množina může disponovat snímky s kvalitou obtížně dosažitelnou na autentických datech, což limituje praktickou využitelnost.

1.2.2 Digitalizace snímků

Digitalizace je proces, při kterém dochází k převodu analogového signálu na signál digitální. K tomuto úkonu lze využít například skener [21]. Stěžejní je volba vhodného rozlišení snímku. V důsledku příliš nízkého rozlišení může dojít k potlačení významných příznaků grafému. Současně je však poměrně náročné dosáhnout vysokého rozlišení, jestliže rukopisy nevznikly za kontrolovaných podmínek.

Zde je potřebné zmínit problematiku zpracování online a offline psaného textu. V druhém případě vlivem digitalizace dochází ke ztrátě dynamických informací o písmu, které jsou pro experty velmi cenné. Mezi ně patří např. rychlost, zrychlení, tlak, sklon psacího nástroje [22].

1.2.3 Předzpracování dat

Jedná se o soubor metod, jejichž účelem je zvýšení kvality snímků, zvýraznění významných příznaků, rozšíření datové množiny. Předzpracování dat významně ovlivňuje výsledky algoritmů [8]. Následuje výpis standardně používaných technik.

Normalizace

Účelem normalizace je vytvoření jednotnosti napříč datovou množinou. **Změna velikosti** se používá za účelem standardizace rozměrů vstupních snímků. Ty bývají často vyžadovány některými metodami v rámci dalšího zpracování, například neuronovými sítěmi. Současně platí, že volba nevhodných rozměrů snímku může taktéž negativně ovlivnit jejich efektivnost. **Detekce náklonu** se uplatňuje v rámci mitigace chyb způsobených například při skenování. Současně se však může jednat o projev variability pisatele. Přílišná redukce tohoto jevu by tedy mohla vést ke ztrátě významných příznaků. Tato technika je vhodná spíše pro úlohy zabývající se rozpoznáváním znaků [19]. Pro účely verifikace je vhodná například centralizace daného písmene.

Segmentace

Úkolem segmentace je rozdělení snímku na specifické oblasti. Jednou z technik je prahování, jehož výsledkem je binarizace vstupu. Základním principem je komparace jednotlivých hodnot pixelů s prahovou hodnotou. Jestliže je hodnota pixelu menší než prahová, pak je nastavena na minimální možnou hodnotu, v opačném případě na hodnotu maximální. Jednou z konkrétních implementací použitých zde [22] je Otsuovo prahování. Tato technika využívá analýzy histogramu za účelem nalezení optimální prahové hodnoty [21].

Redukce šumu

Přítomnost šumu ve snímku se může projevovat nežádoucími změnami jasu či barev, což může vést k následné degradaci kvality [23]. Na jeho zpracování se využívají filtry sloužící k různým účelům. **Gaussův filtr** je zástupcem dolní propusti a důsledkem jeho aplikace je rozostření obrazu. Vychází z konceptu Gaussovy funkce. **Mediánový filtr** je vhodný pro odstranění šumu typu „sůl a pepř“, tedy náhodně umístěné bílé a černé pixely. Ty jsou nahrazeny hodnotou mediánu jasu sousedních pixelů. **Sobelův filtr** detekuje hrany na základě výpočtu gradientu jasu pixelů v horizontálním a vertikálním směru [21].

Augmentace dat

Augmentace dat představuje možné řešení problému malé trénovací množiny, který je již zmíněn zde 1.2.1. Rotace, zkosení či posun snímku vytváří běžné odchylky v rukopise způsobené například změnou úhlu psaní či tlaku na papír. Přiblížení, či oddálení mohou teoreticky pomoci modelu při identifikaci specifických příznaků. Změnu tloušťky písma v důsledku proměnlivosti tlaku na papír lze potenciálně uměle vytvořit pomocí morfologických operací, jako je dilatace nebo eroze [21, 19]. Elastickou deformaci lze využít pro simulaci variability plynoucí z mimovolných pohybů svalů ruky [24].

1.2.4 Extrahování příznaků a klasifikace

Identifikaci důležitých atributů dat lze provést prostřednictvím extrahování příznaků. Netýká se to pouze zpracování obrazu. Například moderní přístup pro detekci malware taktéž využívá tohoto mechanismu [25]. Výstupem klasifikace je rozhodnutí, zda autorem referenčního a sporného materiálu je tentýž člověk, či nikoliv. Oba tyto kroky jsou velmi specifické. Odvíjí se od konkrétně zvolených přístupů. Detailně jsou rozebrány jak z pohledu konvenčních, tak moderních přístupů v kapitole 1.3 a 1.4.

1.2.5 Zhodnocení algoritmů

Evaluační metriky slouží k ohodnocení spolehlivosti využitých metod. Zde uvedené metriky jsou vybrány na základě binární povahy rozhodovacího problému, kterým se tato práce zabývá.

Preciznost (*precision*) udává podíl správně označených pozitivních případů z celkového počtu predikovaných pozitivních případů, přičemž pro výpočet lze využít následující vztah:

$$\text{preciznost} = \frac{n_{\text{TP}}}{n_{\text{TP}} + n_{\text{FP}}}, \quad (1.1)$$

kde n_{TP} je počet pravdivých pozitivních predikcí, n_{FP} je počet falešných pozitivních predikcí [26].

Míra pravdivých pozitivních predikcí (*True Positive Rate, TPR*), známá také jako **senzitivita** či **recall**, vyjadřuje poměr správně klasifikovaných pozitivních případů ke všem skutečně pozitivním případům. Jinak řečeno, měří schopnost modelu správně identifikovat pozitivní případy a lze ji vyjádřit jako

$$\text{TPR} = \frac{n_{\text{TP}}}{n_{\text{TP}} + n_{\text{FN}}}, \quad (1.2)$$

kde n_{TP} je počet pravdivých pozitivních predikcí, n_{FN} je počet falešných negativních predikcí [26].

Míra pravdivých negativních predikcí (*True Negative Rate, TNR*) vyjadřuje schopnost modelu správně identifikovat negativní případy. Tento ukazatel se také označuje jako **specifita** a lze vypočítat jako

$$TNR = \frac{n_{TN}}{n_{TN} + n_{FP}}, \quad (1.3)$$

kde n_{TN} je počet pravdivých negativních predikcí, n_{FP} je počet falešných pozitivních predikcí [26].

Přesnost (*accuracy*) je definována jako poměr všech správně určených případů k celkovému počtu případů. Za hlavní výhodu lze považovat její jednoduchost a intuitivní chápání. Tato metrika je však nevhodná pro nevyvážené datové množiny, kde může poskytovat zavádějící, tj. příliš optimistické výsledky [27]. Její hodnotu lze získat prostřednictvím vztahu:

$$\text{přesnost} = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{TN} + n_{FP} + n_{FN}}, \quad (1.4)$$

kde n_{TP} je počet pravdivých pozitivních predikcí, n_{TN} je počet pravdivých negativních predikcí, n_{FP} je počet falešných pozitivních predikcí a n_{FN} je počet falešných negativních predikcí [26].

Matice záměn informuje o algoritmem správně a mylně klasifikovaných třídách [26]. Třída v tomto případě reprezentuje potvrzení, či vyvrácení autorství. Výstupy získané z matice záměn poskytují detailnější přehled o schopnostech modelu, tudíž mohou být přínosem při jeho optimalizaci. Řádky této dvoudimenzionální matice znázorňují skutečné třídy, zatímco sloupce odpovídají třídám předpovězeným modelem. Popřípadě může být zvoleno opačné uspořádání. Správně vyhodnocené případy se vždy nachází pouze na hlavní diagonále matice.

F1 skóre představuje harmonický průměr *preciznosti* a *senzitivity*, viz 1.5 [26]. Tuto metriku lze využít i při manipulaci s nevyváženým datasetem. V některých případech však nemusí být zcela vhodná vzhledem k invarientní záměně tříd a skutečnosti, že nebere v potaz TN vzorky [27]. Hodnotu této metriky lze stanovit na základě následujícího výpočtu

$$F1 = 2 \cdot \frac{\text{preciznost} \cdot \text{senzitivita}}{\text{preciznost} + \text{senzitivita}}. \quad (1.5)$$

ROC (Receiver Operating Characteristic) křivka a hodnota **AUC** (*Area Under Curve*) bývá často používána k hodnocení výkonu modelu, zejména v případech binární klasifikace. ROC křivka zobrazuje vztah mezi *TPR* a mírou falešných pozitivních predikcí (*False Positive Rate, FPR*) při různých prahových hodnotách. Hodnota *AUC* reprezentuje obsah plochy pod ROC křivkou, přičemž nabývá hodnot z intervalu [0, 1]. Čím vyšší tato hodnota je, tím lepší rozlišovací schopnosti model má [26].

1.3 Konvenční metody

Za konvenční metody v oblasti expertizy ručního písma jsou považovány ty, jež využívají ručně navržené extraktory příznaků [9, 17]. Jejich podstatou je proces, kdy odborník předem definuje, které charakteristiky písma budou analyzovány a jakým způsobem. Některé přístupy se zaměřují na geometrické vlastnosti grafému, jako je výška, šířka, zakončení tahu. Pro popis charakteristických příznaků rukopisu lze využít model směsi Gaussovských rozložení (Gaussian Mixture Model, GMM) [9, 17].

K detekci klíčových bodů v obraze, které jsou invariantní vůči změně měřítka či rotaci, je určen algoritmus SIFT (Scale-Invariant Feature Transform), který lze popsat následujícím způsobem. Nejprve jsou vstupní snímky postupně rozostřovány prostřednictvím konvoluce s gaussovskou funkcí. Následně jsou vytvořeny snímky, které vznikly jako rozdíl různě rozostřených verzí obrazu. Lokální extrémny jsou detekovány porovnáváním hodnot pixelů. Tyto extrémny jsou považovány za potenciální klíčové body, přičemž jsou následně eliminovány ty, které mají například nízký kontrast. Klíčovým bodům je přiřazena orientace dle směřování gradientů v jejich okolí. Na závěr jsou stanoveny deskriptory klíčových bodů, jež jsou využity při komparaci snímků [28].

Poměrně známým algoritmem pro extrahování příznaků je LBP. Je založen na porovnávání hodnot jasů jednotlivých pixelů v předem specifikované oblasti. Každému pixelu je na základě této komparace přiřazena binární hodnota podle toho, zda je jeho jas vyšší než jas centrálního pixelu, či nikoli. Takto získané binární číslo je převedeno do desítkové soustavy, přičemž vyjadřuje prostorovou strukturu dané části snímku. V závěrečné fázi jsou tyto hodnoty využity při tvorbě histogramů, které slouží k charakterizaci snímku [29]. Tyto algoritmy jsou často využívány v kombinaci s klasifikátory, jakými jsou k -NN nebo SVM. Dalším příkladem konvenčního klasifikátoru je Skrytý Markovův model (Hidden Markov Model, HMM) [17].

Ne vždy však tyto přístupy dosahují požadovaných výsledků [19, 30]. Za jejich nedostatek lze považovat závislost na lidském faktoru při určování zkoumaných charakteristik písma. Zdroj [13] klasifikuje jako tradiční taktéž přístupy na bázi strojového učení využívající mělké neruonové sítě (Shallow Neural Network, SNN). Proces expertizy je v tomto případě rozdělen do dvou separátních fází – extrahování příznaků a závěrečná klasifikace. Nedostatkem je vysoká citlivost na kvalitu příznaků ve vstupních datech, kterou je u rukopisů často obtížné zajistit. Tyto příznaky mají klíčovou roli při verifikaci autora, tudíž je vhodnější aplikovat robustnější metody. Inovativnější přístup k řešení této problematiky představují metody na bázi hlubokého učení, jež v posledních letech nabývají na popularitě [9].

1.4 Moderní metody

Moderní metody jsou založeny na principech hlubokého učení. Jedná se o užší oblast strojového učení zaměřenou na architektury s četným množstvím vrstev neuronů, které svým uspořádáním vzdáleně připomínají strukturu mozku savců [31]. Na rozdíl od ručně navržených extraktorů jsou koncipovány tak, aby se samostatně naučily identifikovat významné příznaky písma [9]. Tento přístup také umožňuje plně automatizovat proces analýzy, jelikož model zvládne extrahovat příznaky a provést následnou klasifikaci v rámci jedné ucelené architektury. Ve srovnání s SNN jsou hluboké neuronové sítě (Deep Neural Network, DNN) schopny lépe pracovat se surovými daty, přičemž dosahují lepších výsledků. Mezi nevýhody patří například vyšší nároky na hardware komponenty či v některých případech požadavky na rozsáhlost datových množin [13]. Pro lepší pochopení popisované problematiky je nezbytné definovat základní pojmy, jako jsou umělý neuron (dále jen neuron), parametry, učení a propagace.

Neurony jsou elementární stavební jednotky, jejichž funkci lze popsat následovně. Nejprve je násobením jednotlivých vstupů s příslušnou individuální váhou vypočten vážený součet. K výsledné hodnotě se poté přičítá bias, což poskytuje flexibilitu při aktivaci neuronu. Závěrečným krokem je aplikace aktivační funkce. Váhy a biasy se souhrně označují jako parametry. Matematická formulace celého procesu zní:

$$y(k) = F \left(\sum_{i=0}^m w_i(k) \cdot x_i(k) + b \right), \quad (1.6)$$

kde $x_i(k)$ představuje vstup, $w_i(k)$ odpovídá váze, b označuje bias, F je aktivační funkce a $y(k)$ výstupní hodnota, přičemž všechny veličiny jsou v diskrétním čase k . Způsob, jakým tyto elementární jednotky zpracovávají informace, je inspirován činností biologických neuronů, které se skládají z dendritů, somy a axonu. Dendrity akceptují signály od okolních neuronů, soma je zpracovává, axon je následně přenáší k ostatním buňkám [32]. Z principů konekcionismu vyplývá, že jednotlivé neurony samy o sobě příliš přínosné nejsou. Svůj největší potenciál vykazují až v rámci vzájemně propojené sítě [33].

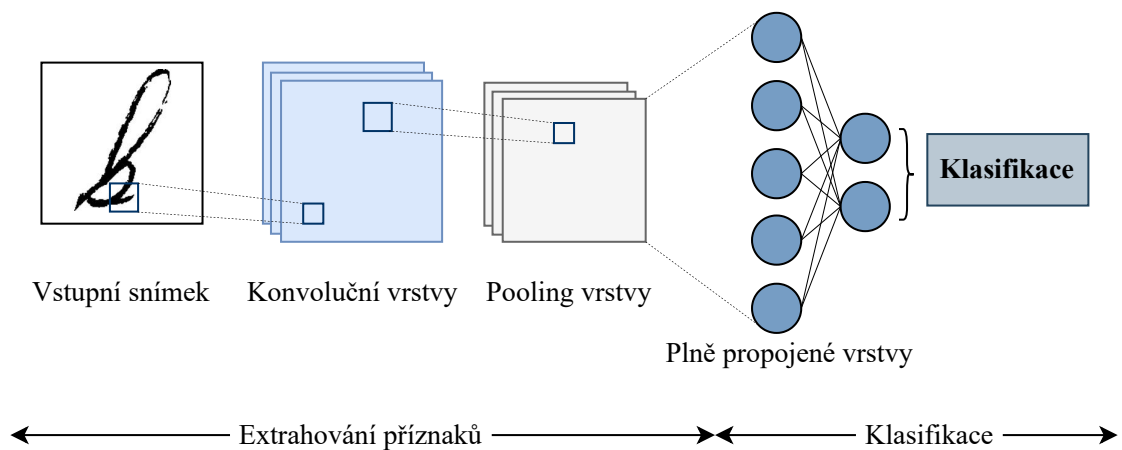
V rámci strojového učení lze rozlišit tři přístupy, kterými jsou učení bez učitele (unsupervised learning), učení s učitelem (supervised learning) a posilované učení (reinforcement learning). Učení bez učitele pracuje s neoznačenými daty a zaměřuje se především na popis vztahů a struktur ve vstupech. Učení s učitelem využívá označená data, na základě kterých se model učí predikovat, do které kategorie daný vstup patří. Třetím přístupem je posilované učení, kde model je penalizován, či odměňován na základě výsledků svých akcí [31]. V rámci této práce bude využito pouze učení s učitelem.

Základním mechanismem, který sítě hlubokého učení využívají k optimalizaci výkonu během fáze trénování, je dopředná a zpětná propagace. Během dopředné propagace model vytvoří predikci, která je následně porovnána se skutečnou hodnotou. Rozdíl mezi těmito dvěma hodnotami se uplatňuje při zpětné propagaci, během které jsou parametry modelu upraveny tak, aby byl tento rozdíl minimální [33].

Následující podkapitoly se věnují významným konceptům v oblasti hlubokého učení, kterými jsou konvoluční neuronové sítě, siamské sítě a transformersy. Tyto architektury vykazují schopnost zachytit i komplexní vzory, což je vzhledem k variabilitě písma zásadní.

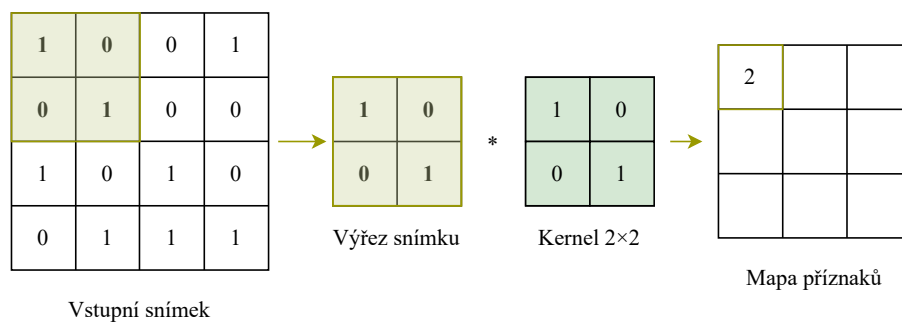
1.4.1 Konvoluční neuronové sítě

Konvoluční neuronové sítě (Convolutional Neural Network, CNN) jsou významným nástrojem v oblasti počítačového vidění [34]. Principiálně vycházejí z mechanismů zrakové percepce živočichů, přičemž klíčové poznatky z této oblasti představili neurofyziologové D. H. Hubel a T. N. Wiesel [35]. Tyto poznatky se projevují v hierarchickém uspořádání CNN. Skládají se z vrstev obsahujících neurony, které jsou zaměřeny pouze na specifické části snímku. Toto uspořádání umožňuje postupně extrahovat příznaky, od jednoduchých linií až po komplexní vzory. Jedná se tedy o zjednodušenou analogii ke zpracování obrazu zrakovým ústrojím a mozkem. Základní komponenty architektury těchto sítí představují konvoluční, podvzorkovací (pooling) a plně propojené vrstvy, dále pak aktivační a ztrátové funkce [34]. Obecné schéma CNN se nachází na obrázku 1.2.



Obr. 1.2: Obecné schéma CNN

Klíčovým principem konvoluční vrstvy je konvoluce.¹ Tato matematická operace slouží k detekci významných příznaků, jakými jsou hrany, křivky a tvary, prostřednictvím filtrů. Ty bývají někdy označovány také jako kernely a jsou postupně aplikovány na jednotlivé části snímku, viz obrázek 1.3. Výstupem celého procesu je mapa příznaků (feature map). Jednotlivé filtry obsahují váhy, jež jsou během trénování modelu optimalizovány tak, aby se zaměřovaly primárně na relevantní informace. Výhodou této vrstvy je, že umožňuje sdílení parametrů. Dochází tak ke snížení paměťových nároků modelu, jelikož stejná sada vah se používá opakovaně na různých pozicích snímku [33, 34]. Konvoluce jsou snadno paralelizovatelné, čímž se zvyšuje rychlost při trénování s využitím grafických procesorů. Na druhou stranu, fixní velikost filtrů omezuje schopnost CNN detekovat příznaky různých měřítek [36].



Obr. 1.3: Operace konvoluce

Podvzorkovací vrstva přispívá k redukci výpočetní složitosti prostřednictvím snížení dimenzionality snímku. Základní princip spočívá v tom, že výstupy z přechodí vrstvy jsou dle konkrétních oblastí agregovány a následně nahrazeny jednou společnou hodnotou. Známé přístupy využívají například substituci maximální (Max Pooling) nebo průměrnou hodnotou (Average Pooling) [33]. Podvzorkovací vrstva se také podílí na mitigaci rizika přetrénování modelu a činí ho robustnějším [34]. Vzhledem k výpočetním možnostem dnešních grafických procesorů nejsou podvzorkovací vrstvy vždy zcela nezbytné.

Plně propojená vrstva se zpravidla nachází na konci CNN. Název vychází ze skutečnosti, že všechny neurony předchozí vrstvy jsou propojeny s těmi z vrstvy aktuální. Přijímá doposud extrahované příznaky a může být následována výstupní vrstvou, kde je prostřednictvím příslušné aktivační funkce provedena finální klasifikace. Příznaky jsou zpracovávány na globální úrovni, což ji činí rozdílnou od konvoluční a podvzorkovací vrstvy [34].

¹Knihovny strojového učení často implementují křížovou korelaci, ale z konvence ji nazývají konvolucí, protože je funkčně podobná.

Aktivační funkce se typicky aplikuje na výstupy konvoluční a plně propojené vrstvy. Do modelu vnáší nelinearitu, což mu umožňuje se lépe adaptovat na problémy z reálného světa. Jednou z nejnámějších aktivačních funkcí je ReLU (Rectified Linear Unit), která ponechává kladné vstupní hodnoty nepozměněné, zatímco záporné hodnoty nastavuje na 0. Výhodou ReLU je, že zmírňuje riziko mizejícího gradientu (vanishing gradient), na rozdíl od funkce sigmoid [34]. Dalšími příklady jsou softmax, vhodný pro vícetřídní klasifikační problémy, či Parametric ReLU [33].

O schopnosti modelu správně klasifikovat data, vypovídá ztrátová funkce $\mathcal{L}[\phi]$. Cílem je nalézt optimální parametry jednotlivých vrstev $\hat{\phi}$ za účelem minimalizace její hodnoty, což lze obecně vyjádřit jako

$$\hat{\phi} = \arg \min_{\phi} [\mathcal{L}[\phi]]. \quad (1.7)$$

Pro dosažení tohoto stavu lze využít gradientní sestup. Nejprve jsou heuristicky stanoveny počáteční hodnoty parametrů $\phi = [\phi_0, \phi_1, \dots, \phi_N]^T$. Poté se vypočítají parciální derivace ztrátové funkce podle jednotlivých parametrů:

$$\frac{\partial \mathcal{L}}{\partial \phi} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \phi_0} \\ \frac{\partial \mathcal{L}}{\partial \phi_1} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial \phi_N} \end{bmatrix}, \quad (1.8)$$

jejichž hodnoty umožňují určit směr jejího poklesu. Parametry jsou následně aktualizovány tak, aby se funkce postupně přibližovala svému minimu, a to dle vzorce:

$$\phi \leftarrow \phi - \alpha \cdot \frac{\partial \mathcal{L}}{\partial \phi}, \quad (1.9)$$

kde α představuje míru učení (learning rate). Celý proces probíhá iterativně. Existují různé variace gradientního sestupu, například Adam (Adaptive moment estimation) nebo stochastický gradientní sestup [31]. Známou ztrátovou funkcí je binární křížová entropie, kterou lze vyjádřit jako

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (1.10)$$

kde N označuje celkový počet vzorků v datové množině, y_i vyjadřuje skutečnou příslušnost vzorku ve třídě a \hat{y}_i predikovanou příslušnost vzorku ve třídě [31].

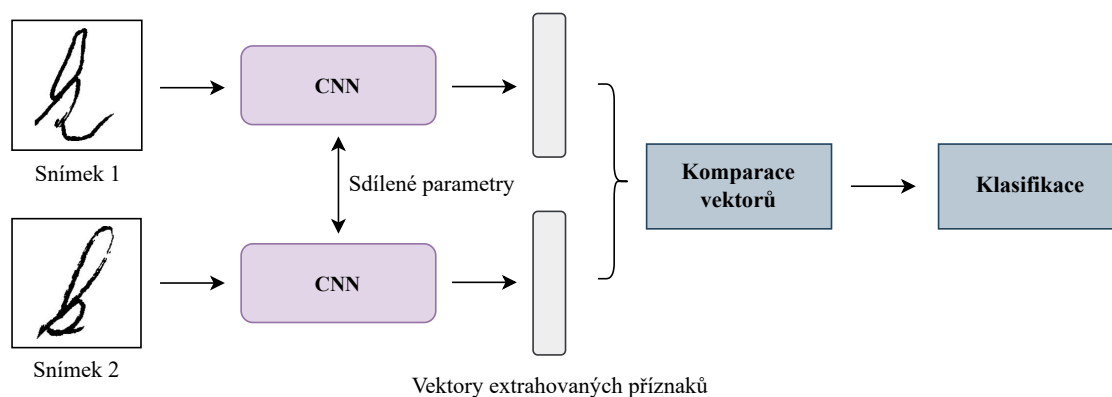
Ačkoliv CNN představují významný typ architektur ve sféře hlubokého učení, mají i své nedostatky. Jedním z nich je absence interpretovatelnosti, tyto modely bývají někdy popisovány jako tzv. black box. Dále citlivost na konfiguraci hodnot jednotlivých parametrů, což je poměrně náročná disciplína. S rostoucí komplexností sítě se pojí riziko přetrénování, vyšší chybovost a paměťové nároky. Mezi

podpůrné nástroje, jež se využívají k minimalizaci těchto negativních vlivů patří například implementace teplotních map (heat maps). Ty poskytují vizuální reprezentaci výstupů. Vynechávající vrstva (dropout layer), normalizace dávek (batch normalization), L1 či L2 regularizace se využívají při řešení problematiky spojené s přetrénováním [34]. Časovou a paměťovou úsporu v důsledku snížení počtu parametrů poskytují metody prořezávání (network pruning) [37]. Zmenšení nároků na výpočetní zdroje přináší tzv. destilace znalostí (knowledge distillation) založená na přenosu znalostí z komplexního modelu učitel (teacher model) na jednodušší model student (student model) [38].

Potenciál vykazují sítě využívající konvoluce i při řešení problematiky expertizy ručního písma, viz zdroj [17]. Studie [39] dosahuje poměrně vysoké *přesnosti* při verifikaci podpisů prostřednictvím architektur na bázi VGG a ResNet. Taktéž se jeví jako vhodný nástroj v oblasti detekce falzifikátů, jelikož jsou schopny určit padělky i na základě minimálních odchylek [5].

1.4.2 Siamské sítě

Siamské sítě se skládají ze dvou identických neuronových podsítí. Každá z nich zpracovává vstup samostatně a následně generuje vektor extrahovaných příznaků. Tato architektura je navržena tak, aby umožnila porovnání dvojic snímků a stanovení míry jejich podobnosti [40]. Z důvodu konzistence a rovnocenných podmínek při komparaci jsou parametry v podsítích shodné. Základní schéma architektury se nachází na obrázku 1.4.



Obr. 1.4: Obecné schéma siamské sítě

Fáze učení siamské sítě spočívá v optimalizaci schopnosti modelu rozlišovat mezi vstupními snímky. Získané vektory extrahovaných příznaků jsou zpracovány prostřednictvím metriky vzdálenosti, jejíž výsledek vypovídá o tom, zda jsou vstupy podobné, či rozdílné. Běžně používanou metrikou je euklidovská vzdálenost [41]. Při

tréninku modelu se obvykle používá kontrastivní ztrátová funkce (contrastive loss), která penalizuje chyby při klasifikaci podobných a rozdílných párů. Lze ji vyjádřit jako

$$\mathcal{L}_c = \frac{1}{2N} \sum_{i=1}^N \left(y_i \cdot d_i^2 + (1 - y_i) \cdot \max(0, m - d_i)^2 \right), \quad (1.11)$$

kde d_i je vzdálenost mezi vektory extrahovaných příznaků, y_i nabývá hodnoty 1 pro páry podobné a hodnoty 0 pro páry rozdílné, m je předem nastavená prahová hodnota (margin) určující minimální požadovanou vzdálenost mezi rozdílnými dvojicemi [42]. Siamská síť může být rozšířena na tři podsítě, kde vstupem je kotva (anchor), pozitivní případ (rukopis stejného autora) a negativní případ (rukopis jiného autora). Pro tuto architekturu je vhodné použít tripletovou ztrátovou funkci (triplet loss), definovanou následovně:

$$\mathcal{L}_{\text{triplet}} = \frac{1}{N} \sum_{i=1}^N \max \left(d_{(a,p)}^{(i)} - d_{(a,n)}^{(i)} + m, 0 \right), \quad (1.12)$$

přičemž $d_{(a,p)}^{(i)}$ je vzdálenost mezi kotvou a pozitivním případem a $d_{(a,n)}^{(i)}$ je vzdálenost mezi kotvou a negativním případem [42].

Výhodou siamských sítí je, že vykazují poměrně dobré výsledky i při nedostatku dat [41]. Což může být pro tuto práci přínosné vzhledem k úbytku psaného projevu. Dále také bývají využívány v oblasti expertizy ručního písma, například při identifikaci autora či verifikaci podpisu, přičemž jsou modifikovány tak, aby byly schopny zachytit i nuance rukopisu [40, 43]. Dle článku [11] lze tyto architektury považovat za obzvláště vhodné při řešení této problematiky.

1.4.3 Transformers

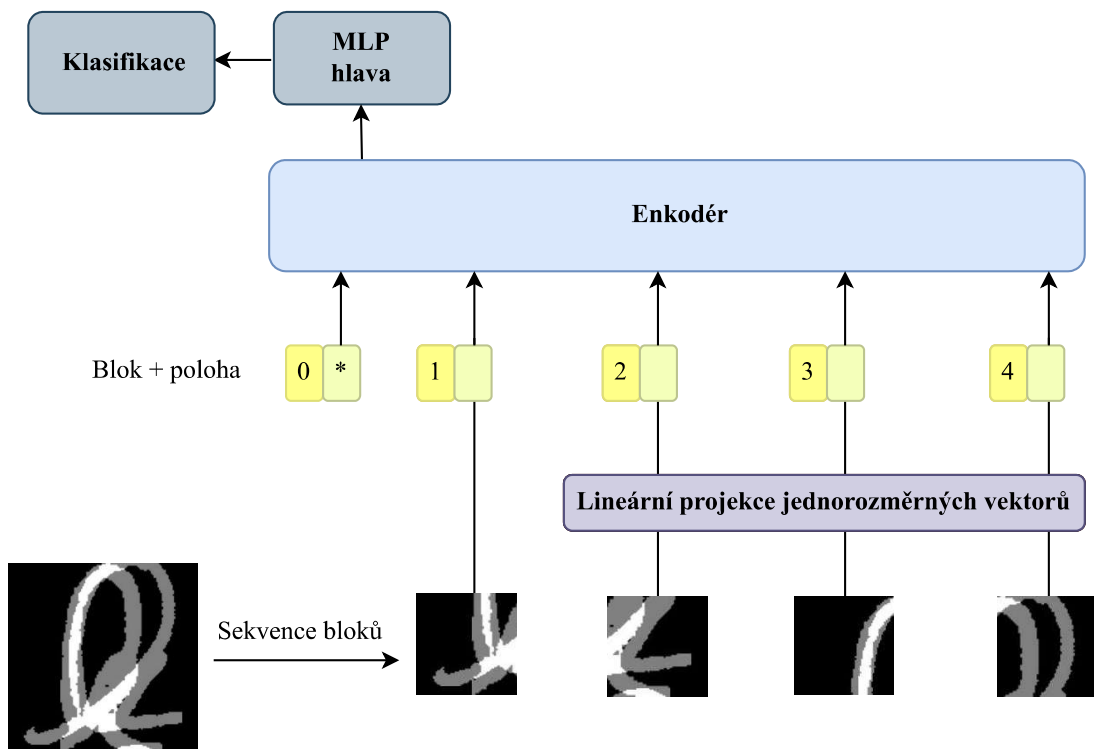
Transformery reprezentují architekturu hlubokého učení, která byla původně cílena na problematiku zpracování přirozeného jazyka [31]. Vynikají především schopností efektivně pracovat se sekvenčními daty většího rozsahu. Mezi známé příklady patří modely GPT. Ačkoliv to nemusí být na první pohled patrné, existuje jistá podobnost mezi datovými množinami přirozeného jazyka a obrazových dat. V obou případech může být počet vstupních hodnot velmi vysoký a na různých pozicích se mohou vyskytovat podobné příznaky a charakteristiky [31]. V roce 2021 byl poprvé představen vision transformer (ViT) [44], což je architektura specializovaná na zpracování obrazu, viz obrázek 1.5.

Princip ViT je založen na rozdělení vstupního obrazu na menší bloky stejné velikosti, tzv. patche. Vzniklé sekvence bloků jsou převedeny do jednorozměrných vektorů a ke sjednocení dimenzionality je použita trénovatelná lineární projekce (trainable linear projection). Následně je přidělen token, jenž postupně zaznamenává

informace o jednotlivých blocích. Důležitou součástí procesu je také přidání informace o pozici jednotlivých bloků, aby model mohl rozpoznávat jejich prostorové uspořádání. Stěžejní komponentou ViT je enkodér obsahující normalizační vrstvu, vícehlavou pozornost (Multi-Head Self-Attention, MHSA) a vícevrstvý perceptron (Multi-Layer Perceptron, MLP) [44]. Díky MHSA je model schopen lépe zachytit významné příznaky vstupních dat tím, že mechanismus pozornosti analyzuje různé oblasti snímku nezávisle na sobě. Tyto dílčí výstupy jsou následně sloučeny do jedné reprezentace. Mechanismus pozornosti lze zapsat rovnicí

$$\text{pozornost}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V, \quad (1.13)$$

kde Q, K, V představují dotaz (query), klíč (key), hodnotu (value) a d_k je dimenze klíčů a dotazů [45]. Účelem funkce softmax je stanovit podobnost mezi dvěma prvky sekvence, a to dle hodnot Q a K . Výstupem této operace jsou váhy pozornosti (attention weights), jež jsou následně násobeny s V . V závěrečné části dochází ke klasifikaci, kterou uskutečňuje MLP na základě hodnoty tokenu [44].



Obr. 1.5: Architektura ViT

Způsob, jakým CNN a ViT zpracovává data, je značně odlišný. Architektura CNN se zaměřuje primárně na lokální příznaky vstupního snímku, jež extrahuje prostřednictvím hierarchicky uspořádaných vrstev. Naproti tomu ViT využívá mechanismu

pozornosti za účelem detekce významných závislostí a příznaků z hlediska globálního kontextu vstupu. Na rozdíl od CNN jsou ViT schopny lépe zachytit dlouhodobé závislosti ve vstupní sekvenci dat, avšak za cenu vyšší výpočetní složitosti [36].

Nevýhodou této architektury založené na transformeru jsou vysoké nároky na množství trénovacích dat, což představuje zásadní problém vzhledem k omezenému počtu vzorků grafémů, se kterým se expertiza ručního písma potýká. Proto vznikla odlehčená varianta ViT Lite. Dle studie [46], zabývající se forenzní identifikací autora rukopisu, má potenciál poskytovat výsledky odpovídající současnému stavu vědy. Zdroj [47] nabízí srovnání využití ViT a CNN architektur ResNet, EfficientNet pro porovnání dvou ručně psaných dokumentů, přičemž ViT v tomto případě vykazoval nejlepší výsledky. Dalším přístup spočívá v kombinaci výše uvedených typů architektur, například propojení konvolučních vrstev a mechanismu pozornosti v rámci siamské sítě [48].

1.5 Přehled známých architektur

Při řešení problematiky expertizy ručního písma pomocí umělé inteligence lze vytvořit vlastní modely, nebo využít již existující, které lze případně rozšířit. Tato kapitola se zabývá popisem známých architektur, které bývají implementovány ve studiích zabývajících se touto problematikou. Výhodou přístupu využívajícího přenosové učení je časová úspora a podpora schopnosti generalizovat [18]. Přínosem tohoto detailního rozboru je pochopení principů, na základě kterých tyto sítě byly vytvořeny a současně se jedná o zdroj inspirace při návrhu vlastních architektur. Níže jsou specifikovány modely VGG, ResNet a EfficientNet, a to na základě zdrojů [13, 17, 18, 47].

1.5.1 Architektura VGG

Model VGG, pojmenovaný dle výzkumné skupiny Visual Geometry Group, vznikl na základě studie zkoumající vliv počtu vrstev sítě na její *přesnost*. Architektura využívá konvoluční kernely velikosti 3×3 , jejichž počet se postupně zdvojnásobuje. Buď 1 nebo až 4 konvoluční vrstvy jsou následovány podvzorkovací vrstvou typu MaxPool. Sít je zakončena třemi plně propojenými vrstvami a funkcí softmax. Použití většího počtu konvolučních vrstev s menšími kernely se ukázalo jako efektivnější ve srovnání s menším počtem vrstev s většími kernely. Hlavními výhodami je snížení počtu parametrů a implementace většího počtu aktivačních funkcí ReLU. Z dosažených výsledků vyplývá, že nižší chybovostí disponuje model s četnějšími vrstvami. Přestože VGG byl původně trénován na datové množině ImageNet, dobrých výsledků dosáhl i pro ostatní datové množiny, čímž byla prokázána schopnost generalizovat [49].

1.5.2 Architektura ResNet

Mezi další zástupce známých architektur patří ResNet, tzv. reziduální síť (residual network). Degradace je jedním z problémů, se kterým se modely hlubokého učení potýkají.² Projevuje se vyšší trénovací chybou, než jaké by bylo dosaženo při použití menšího počtu vrstev. Z toho důvodu ResNet využívá reziduálního učení založeného na přemostění (shortcut connections), jež nezatěžují architekturu dalšími parametry a nezvyšují výpočetní složitost. Přemostění umožňují přímý přenos informací mezi více vrstvami. Původní mapování (underlying mapping) je tak nahrazeno mapováním reziduálním (residual mapping). Matematicky lze tento proces definovat jako

$$y = F(x, \{W_i\}) + x, \quad (1.14)$$

kde y a x je výstupní a vstupní vektor a $F(x, \{W_i\})$ je funkce popisující reziduální mapování. Základní kostra architektury je inspirována VGG. Nejlepších výsledků bylo dosaženo při využití 152 vrstev a tento přístup vykazuje potenciál i při implementaci více jak 1 000 vrstev. Architektura ResNet byla schopna předčít VGG a umístila se na 1. místě v soutěži ILSVRC 2015 [50].

1.5.3 Architektura EfficientNet

Architektura EfficientNet se zaměřuje na škálovatelnost CNN z hlediska její hloubky, šířky a rozlišení snímku, přičemž jsou analyzovány následné dopady na výkonnost modelu. Dle metody složeného škálování (compound scaling) je stěžejní modifikovat tyto tři hodnoty v konstantním poměru za účelem zvýšení *přesnosti* a efektivnosti modelu. Empiricky byla stanovena jejich provázanost, proto je pro tyto účely využit tzv. složený koeficient ϕ , který je patřičně aplikován na všechny tři hodnoty. Základní stavební jednotkou modelu je mobilní invertovaný blok úzkého hrdla (Mobile inverted Bottleneck Convolution, MBConv) s tzv. kompresní a excitační (squeeze and excitation, SE) optimalizací. Aplikace složeného škálování zvýšila *přesnost* i u architektur jako ResNet nebo MobileNet. Účinnost tohoto přístupu se projevila i na tzv. heat mapách analyzovaných snímků z datové množiny ImageNet. Benefitem je nárůst *přesnosti* při menším počtu parametrů [51].

²Není však způsobena přetrénováním. Poukazuje na skutečnost, že pro optimalizaci modelu neexistují univerzální pravidla.

1.6 Teplotní mapy

Jedním z hlavních nedostaků přístupů na bázi hlubokého učení je nízká míra interpretovatelnosti. Částečným řešením tohoto problému je vizualizace prostřednictvím teplotních map. Jejich cílem je podpořit věrohodnost dosažených výsledků formou, která je pro lidského jedince intuitivnější. Může tak snáze odhalit nedostatky daného modelu, a tím přispět k jednodušší optimalizaci. Často používané jsou teplotní mapy založené na gradientně váženém mapování aktivace tříd (Gradient-weighted Class Activation Mapping, Grad-CAM). Jedná se o robustnější přístup, který vychází z CAM. Jak již z názvu Grad-CAM vyplývá, teplotní mapy jsou založeny na hodnotách gradientů, získaných dle rovnice:

$$\frac{\partial y^c}{\partial A_{ij}^k}, \quad (1.15)$$

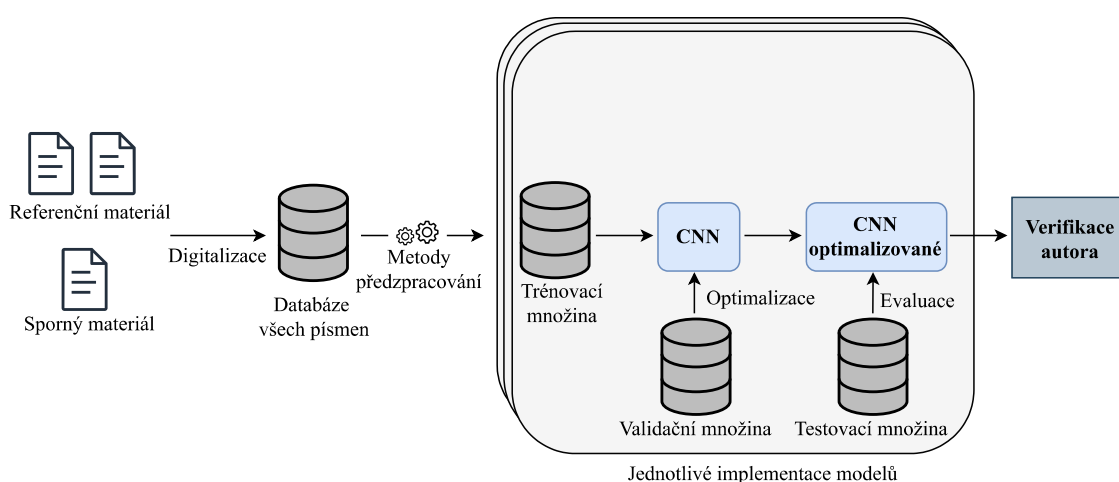
kde y^c je skóre predikované třídy a A_{ij}^k představuje hodnotu na pozici $[i, j]$ v mapě příznaků k . Tyto gradienty jsou následně zprůměrovány přes jednotlivé pozice $[i, j]$, čímž je stanovena významnost dané mapy příznaků vůči predikci. Teplotní mapa je získána dle vzorce:

$$\sum_k \alpha_k^c A^k, \quad (1.16)$$

kde α_k^c reprezentuje významnost mapy příznaků k vůči predikci a A^k označuje mapu příznaků k . V závěru rovněž dochází k eliminaci záporných hodnot, například prostřednictvím aktivační funkce jako je ReLU, či prahováním [52].

2 Praktická část bakalářské práce

Praktická část bakalářské práce se věnuje návrhu a implementaci tří modelů hlubokého učení, jejichž cílem je verifikace autora. Modely jsou založeny na CNN a siamské síti, přičemž volba těchto architektur vychází z poznatků uvedených v teoretické části. Obecné schéma experimentu se nachází na obrázku 2.1. V první fázi dochází ke sběru dat, jež jsou následně předzpracována patřičným způsobem. Dalším krokem je návrh a trénování samotného modelu. Ten je následně ohodnocen evaluačními metrikami. Zvolené přístupy a dosažené výsledky jsou v závěrečné části srovnány s jinými relevantními pracemi zabývajícími se expertizou ručního písma, jež odpovídají současnému stavu vědy.



Obr. 2.1: Obecné schéma experimentu

2.1 Datová množina

Datová množina byla opatřena odborným konzultantem. Vzorky v ní obsažené pochází celkem od 45 osob, přičemž každá z nich poskytla 2–4 strany formátu A4 psaného projevu. Autoři těchto rukopisů pochází z České a Slovenské republiky, v důsledku čehož může jejich písmo vykazovat jistá specifika oproti zahraničním vzorkům. Pro účely digitalizace bylo využito skeneru, přičemž snímky byly skenovány s rozlišením 600 dpi. Následně došlo ke konverzi barevného prostoru do stupňů šedi za účelem mitigace rizika, že se model naučí klasifikovat grafém na základě barvy inkoustu či tuhy. Binarizace obrázku, která je nezbytná pro nadcházející krok, byla provedena prostřednictvím metody prahování. Detekce komponent souvislosti byla uskutečněna funkcí z knihovny OpenCV, jež vychází z teorie grafů. Funkce označuje štítkem s hodnotou 0 všechny pixely tvořící pozadí snímku, zatímco každé souvislé

komponentě přiřazuje unikátní štítek s hodnotou větší než 0 [53]. Následně došlo k manuální selekci písmene a vytvoření konvexní obálky. Výsledné snímky obsahující vybraná písmena byly centralizovány a upraveny na jednotný rozměr 229×229 pixelů. Celkem bylo segmentováno 1 327 znaků „k“ a 3 060 znaků „a“.

Vzhledem k omezenému počtu autorů byl zvolen následující postup při tvorbě datových množin. Pro totožné autory rukopisů byl nejprve náhodně vybrán dokument, který byl označen jako referenční. Následně byl vybrán odlišný dokument téhož autora a označen jako sporný. Poté byly vybrány 2–4 znaky ze sporného a 4–8 znaků z referenčního dokumentu. V případě různých autorů byl postup obdobný, avšak sporný dokument pocházel od jiného náhodně vybraného pisatele. Celkově byly vytvořeny 3 datové množiny: trénovací, testovací a validační. Na trénovacích datech je model učen extrahovat významné příznaky, validační data jsou použita při optimalizaci modelu. Testovací data se využívají pro finální zhodnocení architektury a její schopnosti generalizovat. Počet dat v jednotlivých množinách pro oba modely, rozdělených podle shodného, či rozdílného pisatele, je uveden v tabulce 2.1 pro písmeno „k“ a v tabulce 2.2 pro písmeno „a“. Za účelem ověření robustnosti modelu vůči různým znakům byla vytvořena další trénovací, testovací a validační množina, která vznikla pouhým sloučením odpovídajících množin obou znaků.

Tab. 2.1: Datové množiny pro znak „k“

Trénovací množina		Testovací množina		Validační množina	
Shodný p.	Rozdílný p.	Shodný p.	Rozdílný p.	Shodný p.	Rozdílný p.
421	426	90	90	70	76

Tab. 2.2: Datové množiny pro znak „a“

Trénovací množina		Testovací množina		Validační množina	
Shodný p.	Rozdílný p.	Shodný p.	Rozdílný p.	Shodný p.	Rozdílný p.
372	385	86	88	90	90

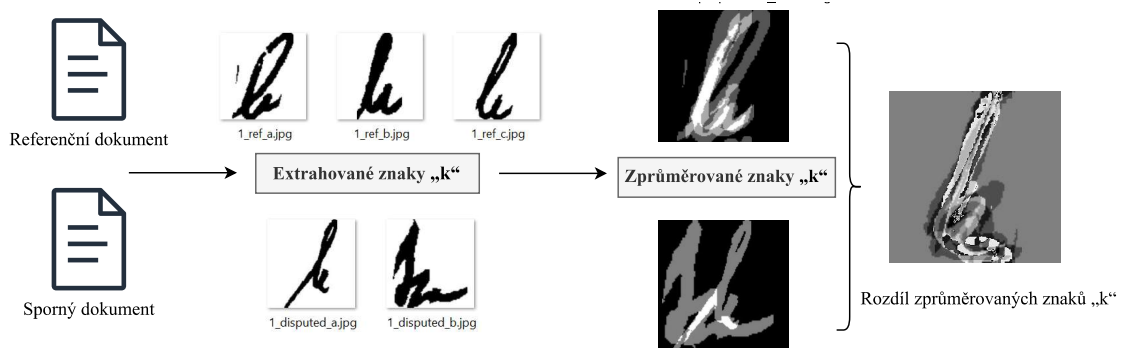
2.2 Navržené metody určení shody

Každá z architektur navržených modelů je zvolena tak, aby odpovídala specifickým požadavkům metod předzpracování dat. V obou případech je trénovací datová množina považována za vyváženou, a proto jsou drobné odchylky eliminovány pomocí numerických metod vyvažování. Samotná implementace je uskutečněna v programovacím jazyce Python. Je zde využita platforma Tensorflow poskytující

API Keras usnadňující vývoj modelů strojového učení. Knihovna CuDNN a nástroj CUDA Toolkit umožňují efektivně akcelarovat výpočty prostřednictvím grafického procesoru. Pro zhodnocení modelů je využito funkcí z knihoven Scikit-learn, OpenCV a Matplotlib.

2.2.1 Model konvoluční neuronové sítě

Tento model využívá základních principů architektury CNN. Vstupem je vždy jeden snímek obsahující rozdíl zprůměrovaných znaků daného písmene, přičemž jejich způsob vzniku je následující. Nejprve jsou zprůměrovány referenční a sporné znaky, které jsou následně vzájemně odečteny. Snímek zobrazuje rozdílné partie v různých stupních šedi. Proces vzniku těchto snímků se nachází na obrázku 2.2.



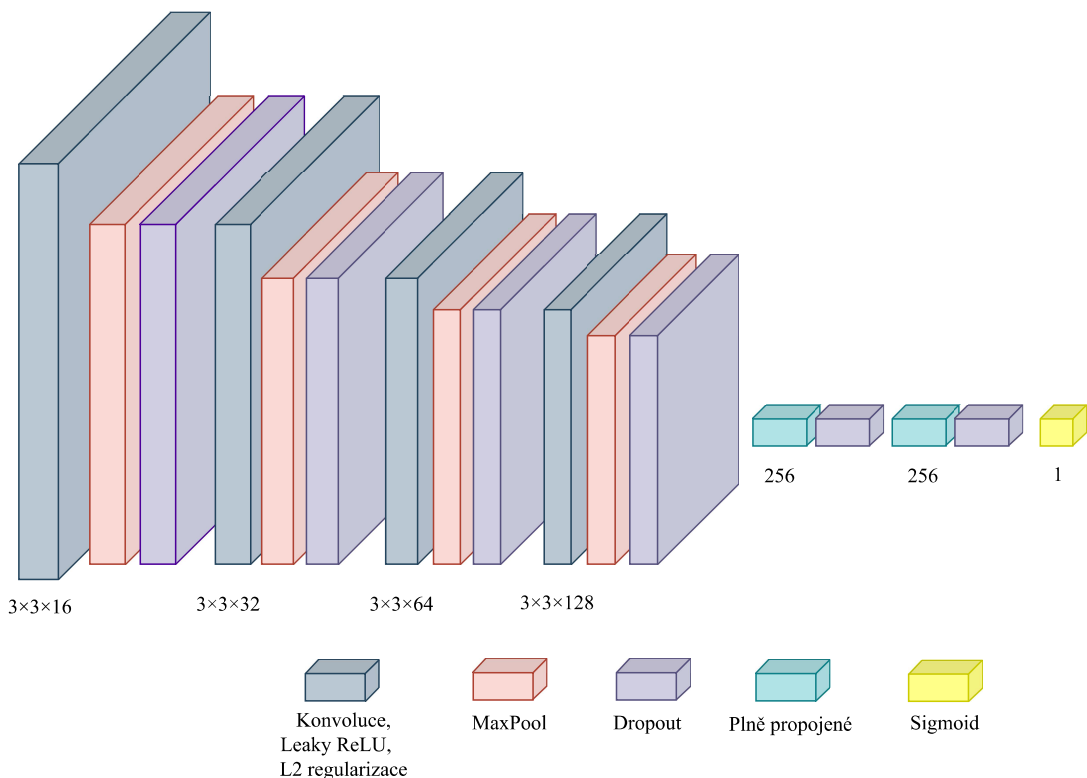
Obr. 2.2: Předzpracování dat pro CNN

Takto předzpracovaným datům je následně přidělen odpovídající štítek (label) reprezentující příslušnost snímku ve třídě. Štítkem hodnoty 0 jsou opatřeny snímky obsahující rozdíl zprůměrovaných znaků od odlišných pisatelů, v opačném případě nese štítek hodnotu 1. Experimentálně bylo zjištěno, že model citlivě reaguje na větší množství augmentačních technik, což může být způsobeno povahou a kvalitou vstupních dat. Z tohoto důvodu byla aplikována pouze rotace, horizontální a vertikální posun, které mohou simulovat variabilitu pisatele a činí model robustnějším. Trénovací data jsou náhodně promíchána, aby se eliminovalo riziko, že model bude závislý na pořadí vstupních dat, nebo že bude preferovat jednu ze tříd.

Architektura se skládá ze čtyř konvolučních vrstev o velikosti kernelu 3×3 . První vrstva obsahuje celkem 16 kernelů, přičemž s každou další vrstvou se jejich počet vždy zdvojnásobí. Součástí každé vrstvy je L2 regularizace, jež penalizuje příliš vysoké hodnoty vah modelu. Menší hodnoty vah vedou k hladší funkci výstupu, což přispívá ke stabilitě modelu a snížení rizika přetrénování [31]. Každá konvoluční vrstva je následována normalizací dávek, aktivační funkcí, podvzorkovací vrstvou

typu MaxPool a vynechávající vrstvou. Účel normalizace dávek spočívá ve standardizaci výstupů jednotlivých vrstev, což podporuje stabilitu dopředné propagace [31]. Jako aktivační funkce je zde použita Leaky ReLU, která na rozdíl od ReLU vykazuje schopnost adekvátně reagovat i na negativní vstupní hodnoty, čímž je řešen problém tzv. dying ReLU [31]. Podvzorkovací vrstva typu MaxPool má rozměr kernelu 2×2 a je následována vynechávající vrstvou, což je opatření, které v tomto případě pomohlo výrazně omezit tendenci modelu přetrénovávat. V závěrečné části modelu se nachází tři plně propojené vrstvy, přičemž poslední z nich provádí klasifikaci prostřednictvím funkce sigmoid.

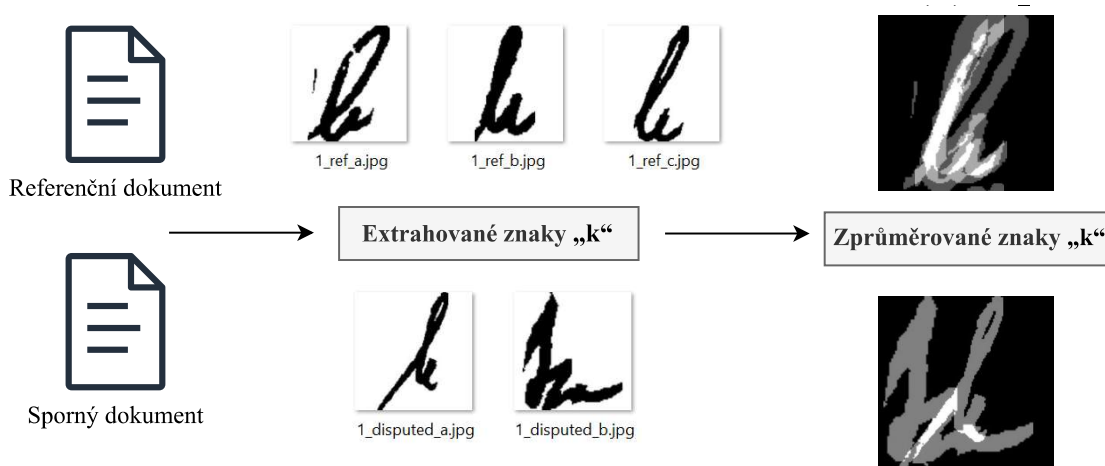
Jako optimalizační metoda gradientního sestupu je použit Adam s výchozí konfigurací parametrů, který adaptivně určuje míru učení α [54]. Ztrátovou funkcí je binární křížová entropie. Dalším prevenčním opatřením před přetrénováním modelu je mechanismus „Early Stopping“, jenž předčasně ukončí proces trénování, jestliže během 10 epoch nedojde k jeho zlepšení. Nástroj „ReduceLROnPlateau“ je nastaven tak, aby snížil míru učení o 20 %, pokud po třech epochách nedojde ke zlepšení hodnot ztrátové funkce. Počet trénovacích epoch je nastaven na 70 a velikost jedné dávky (batch size) je 16. Schéma tohoto modelu se nachází na obrázku 2.3.



Obr. 2.3: Architektura CNN

2.2.2 Model siamské sítě s autoenkodérem

Architektura modelu vychází z konceptu siamské sítě. Vstup tvoří dvojice snímků, přičemž jeden z nich obsahuje zprůměrovaná sporná písmena a ten druhý zprůměrovaná referenční písmena. Tyto snímky vznikly překryvem jednotlivých znaků, které se zobrazují v různých stupních šedi. Vizualizace celého procesu předzpracování dat pro tuto siamskou síť se nachází na obrázku 2.4.

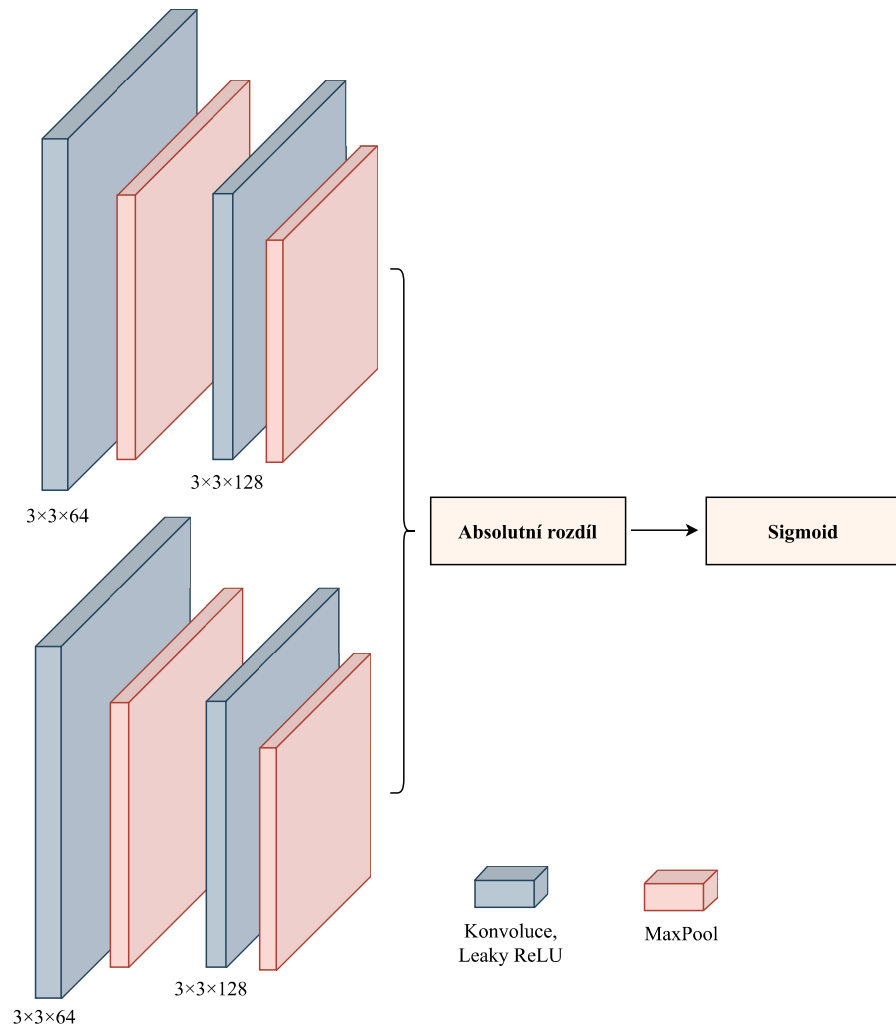


Obr. 2.4: Předzpracování dat pro siamskou síť s autoenkodérem

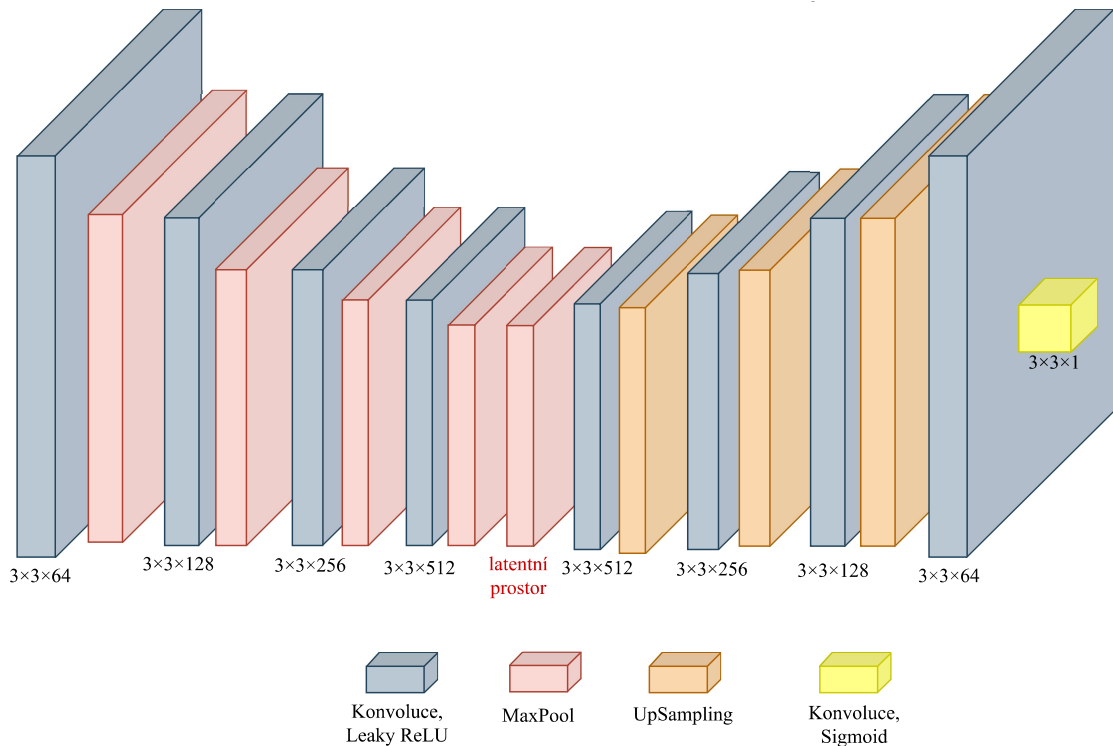
Jednotlivým dvojicím je následně přiřazen štítek. Štítek s hodnotou 0 označuje, že zprůměrované sporné a referenční znaky pochází od totožného autora, zatímco hodnota štítku 1 signalizuje, že pocházejí od autorů různých. Následně je pořadí dvojic náhodně promícháno.

S ohledem na nedostatečnou výkonnost při využití pouze architektury CNN je zde využito autoenkodéru. Jedná se o specifický typ neuronové sítě, který redukuje vstupní data do latentního prostoru a následně se je učí rekonstruovat. Autoenkodér se skládá ze dvou částí: kodéru a dekodéru [31]. Je zde využito konvolučních, a vzorkovacích vrstev typu MaxPool a UpSampling. Kernely jednotlivých konvolučních vrstev mají rozměr 3×3 , přičemž jejich počet se s každou další vrstvou zdvojnásobuje. Jejich rozsah nabývá hodnot 64 až 512. Vzorkovací vrstvy typu MaxPool a UpSampling mají velikost kernelu 2×2 . Jako aktivační funkce je zde použita Leaky ReLU. V poslední vrstvě dekodéru je implementována aktivační funkce sigmoid, jelikož autoenkodér pro optimalizaci parametrů využívá binární křížovou entropii. Schéma autoenkodéru lze vidět na obrázku 2.6. V rámci siamské sítě je využito prvních čtyř vrstev enkodéru, jejichž výstupy jsou následně předány vrstvě Lambda, která určí jejich absolutní rozdíl. Následná klasifikace je vykonána funkcí sigmoid.

Optimalizovaný gradientní sestup je realizována algoritmem Adam. Stejně jako v předchozím případě, i zde je aplikován „Early Stopping“ s počtem epoch nastavených na 10 a „ReduceLROnPlateau“, jenž po 10 epochách snižuje míru učení α o 10 %. Počet trénovacích epoch je 70 a velikost jedné dávky je 16. Schéma modelu siamské sítě se nachází na obrázku 2.5.



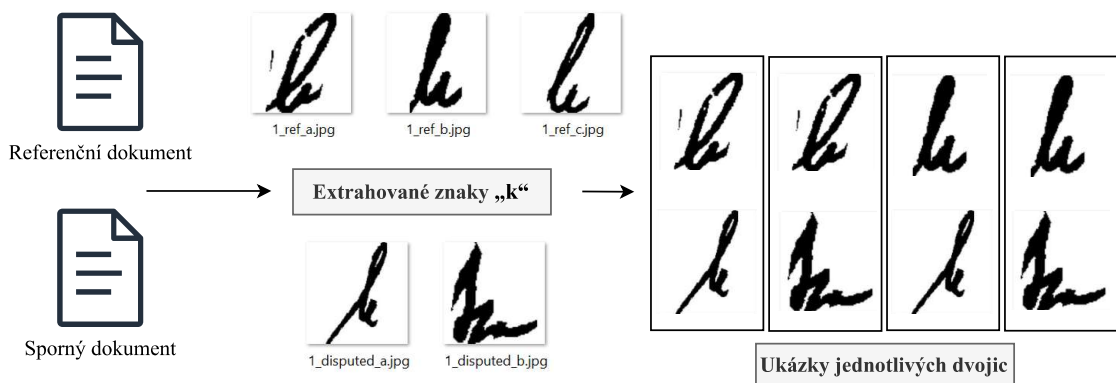
Obr. 2.5: Architektura siamské sítě s autoenkodérem



Obr. 2.6: Autoenkodér – součást siamské sítě

2.2.3 Model konvoluční siamské sítě

Tímto modelem je taktéž siamská síť, jejíž vstupy tvoří snímky obsahující jednotlivé dvojice znaků. Jedná se tedy o nejméně invazivní metodu předzpracování ve srovnání s předchozími metodami. Tento přístup nejlépe imituje způsob, jakým písmoznalec během expertizy postupuje. Proces vzniku jednotlivých dvojic se nachází na obrázku 2.7. Jedná se o výsledek kartézského součinu mezi referenčními a spornými vzorky.

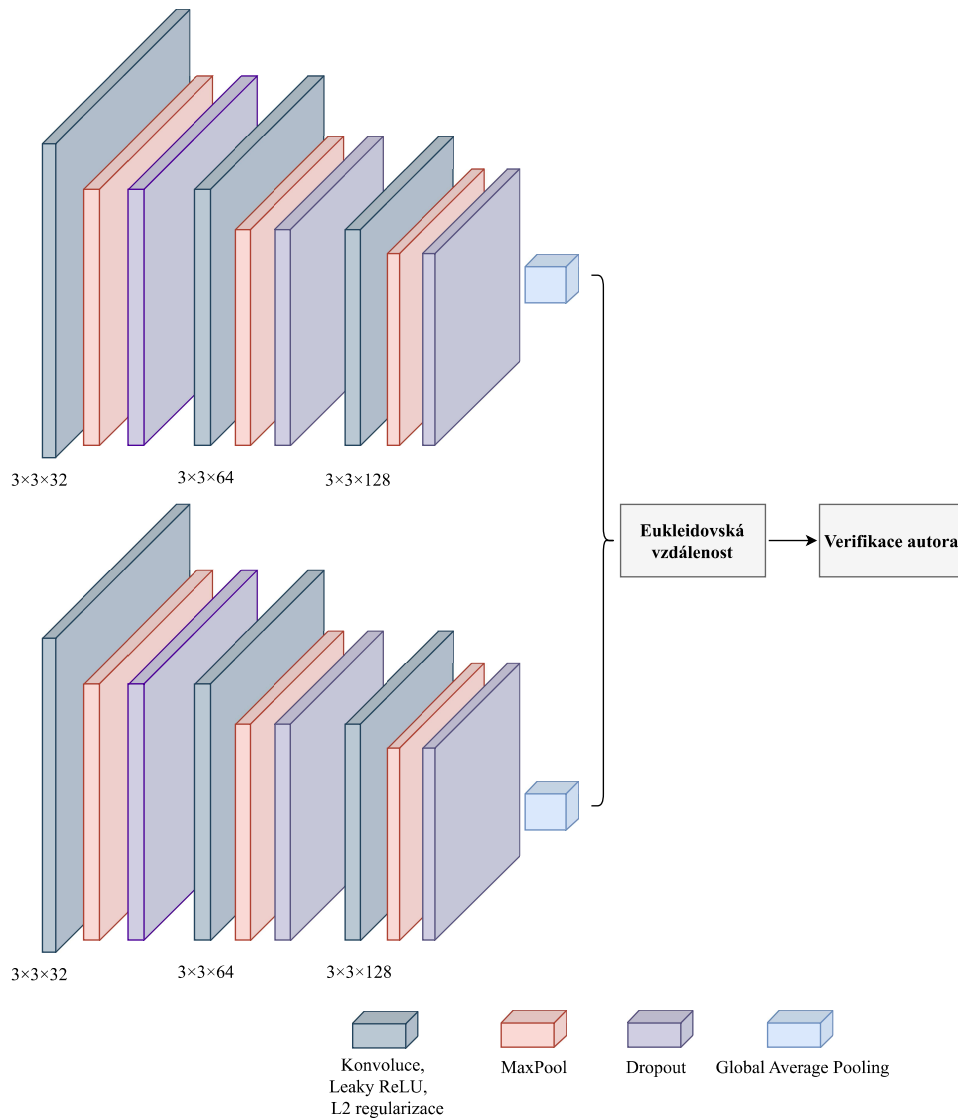


Obr. 2.7: Předzpracování dat pro konvoluční siamskou síť

Dvojice od totožného autora nesou štítek s hodnotu 1 a v opačném případě hodnotu 0. Z důvodu nedostatečné reprezentativnosti již předpřipravené trénovací množiny došlo k jejímu spojení s validační množinou, k promíchání a následnému rozdělení v poměru 8 : 2.

Za účelem extrahování příznaků je zde využito CNN. Ta obsahuje celkem 4 konvoluční vrstvy o rozměru 3×3 , které vždy zdvojnásobují svůj počet kernelů, a to v rozsahu 32 až 128. Ty dále obsahují L2 regularizaci a jsou následovány normalizací dávek. Další regularizační technikou je vynechávající vrstva. Implementované podvzorkovací vrstvy jsou typu MaxPool s velikostí kernelu 2×2 . Poslední vrstvou je podvzorkovací vrstva typu Global Average Pooling, která plní stejnou funkcionalitu jako plně propojená vrstva. Její hlavní výhodou je schopnost zachovat odpovídající vztah mezi mapami příznaků a třídami. Mimo to neobsahuje optimalizovatelné parametry, tudíž není náchylná na přetrénování [55]. Aktivační funkcí je Leaky ReLU. Aktualizace parametrů sítě probíhá na základě výstupů kontrastivní ztrátové funkce s prahovou hodnotou rovnou 2, viz kapitola 1.4.2. Metrikou vzdálenosti mezi extrahovanými vektory z obou podsítí je eukleidovská vzdálenost. Nízké hodnoty této metriky signalizují totožného autora porovnávaných rukopisů. Naopak vysoká hodnota znamená, že se jedná o rozdílné autory. Z těchto důvodů bylo nezbytné invertovat logiku při evaluaci modelu. Klasifikace, čili verifikace autora, je provedena na základě porovnání hodnoty vzdálenosti s prahovou hodnotou. Schéma této architektury zpracovávající jednotlivé dvojice znaků se nachází na obrázku 2.8.

Obdobně jako u předchozích architektur, i zde je za účelem gradientního sestupu využito algoritmu Adam. V rámci další optimalizace je implementován „EarlyStopping“ s počtem epoch 35 a „ReduceLROnPlateau“ snižující míru učení α o 10 % po 5 epochách. Celkový počet trénovacích epoch je 70 a velikost jedné dávky činí 128.



Obr. 2.8: Architektura konvoluční siamské sítě

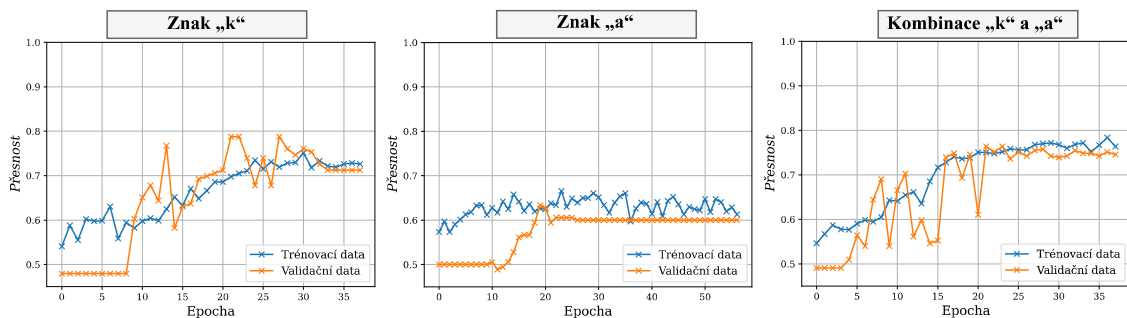
2.3 Dosažené výsledky

Následuje srovnání dosažených výsledků pro všechny tři modely. Vzhledem k binární povaze problému, jímž se tato práce zabývá, a vyváženosti datových množin, jsou zvoleny tyto evaluační metriky: *přesnost*, *preciznost*, *senzitivita*, *F1* skóre, ROC křivka a *AUC*. Vizualizace počtu správně či špatně klasifikovaných snímků se nachází v příslušných maticích záměn. Pro lepší přehled a porozumění jednotlivým modelům jsou uvedeny grafy znázorňující vývoj hodnot *přesnosti* na trénovacích a validačních datech. Všechny uvedené výsledky byly získány na testovacích datových množinách, a to jak pro znak „k“, „a“, tak i pro jejich kombinaci. U všech modelů byly implementovány adaptivní zóny nejistoty, čímž došlo k eliminaci hraničních případů ze závěrečné klasifikace.

2.3.1 Evaluace modelu konvoluční neuronové sítě

Grafy na obrázku 2.9 zachycují průběh trénování CNN. S výjimkou datové množiny písmene „a“ nedochází ke stagnaci *přesnosti* a zjevnému přetrénování modelu. Tabulka 2.3 obsahuje hodnoty jednotlivých metrik. Nejlepších výsledků bylo dosaženo s využitím zóny nejistoty, a to pro datovou množinu „k“. *Přesnost* činí 89,3 %, avšak z matic záměn na obrázku 2.10 vyplývá, že do zóny nejistoty byla zařazena více než 1/3 vzorků. Co se znaku „a“ týče, zde došlo ke značnému poklesu jednotlivých hodnot, a ani dle matice záměn 2.11 model neklasifikoval příliš vyváženě. Důležité však je, že v případě kombinace obou množin se výsledky zásadně neliší ve srovnání s těmi nejlepšími. V tomto případě byl model schopen klasifikovat téměř 2/3 datové množiny s *přesností* 83,8 %, viz 2.12. Z grafů ROC křivek na obrázcích 2.13, 2.14 a 2.15 vyplývá, že model je schopen nejlépe diferencovat mezi shodnými a rozdílnými pisateli v případě znaku „k“, kde hodnota *AUC* je 88,4 %.

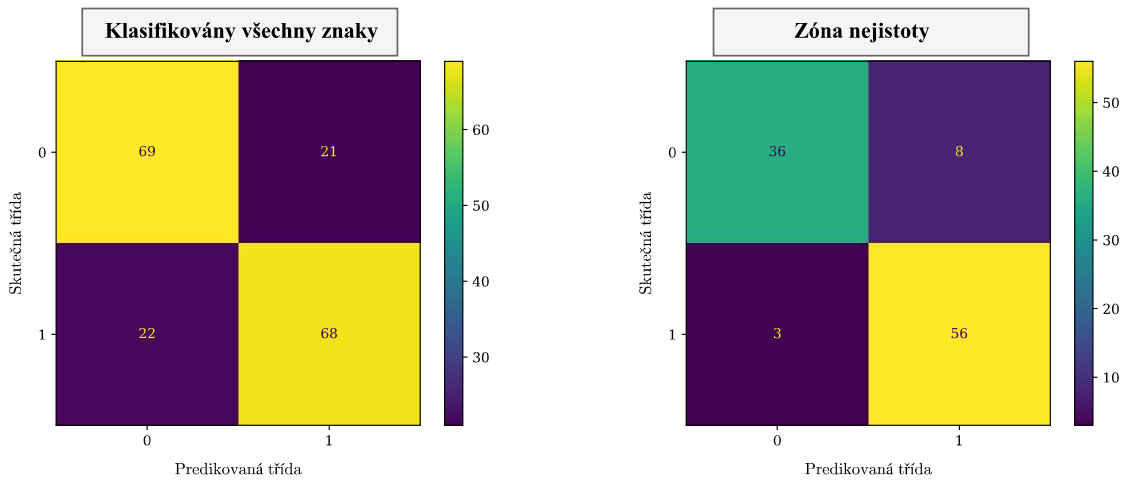
Při implementaci této architektury bylo nejnáročnější dosáhnout schopnosti modelu generalizovat. Síť velmi citlivě reagovala na jednotlivé změny parametrů, tudíž nebylo snadné ji optimalizovat. Byly vyzkoušeny různé modifikace architektury, a to implementace VGG, autoenkodéru, přemostění a mechanismu pozornosti, ty však dosahovaly v nejlepších případech srovnatelných výsledků.



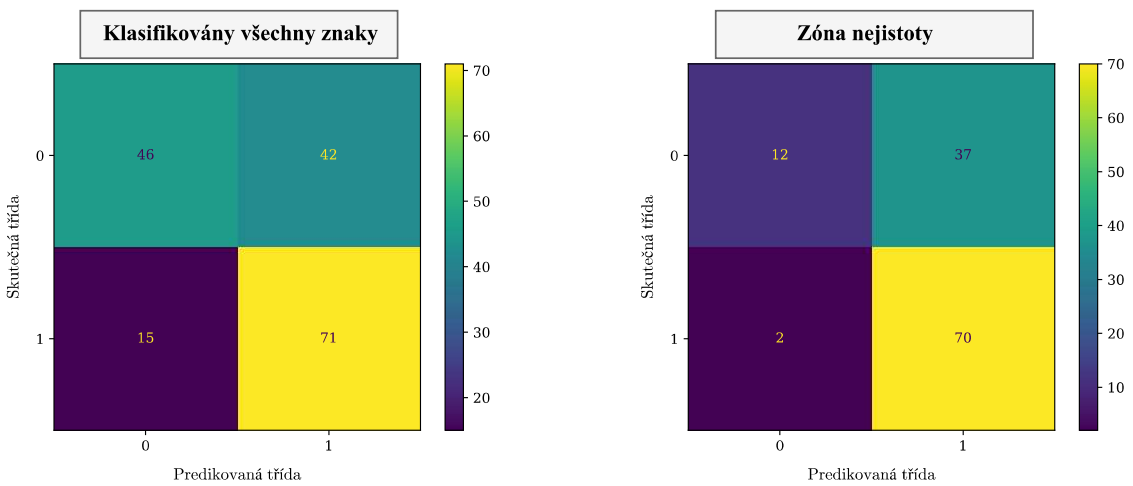
Obr. 2.9: Průběhy jednotlivých trénování CNN

Tab. 2.3: Metriky CNN pro písmena „k“ a „a“

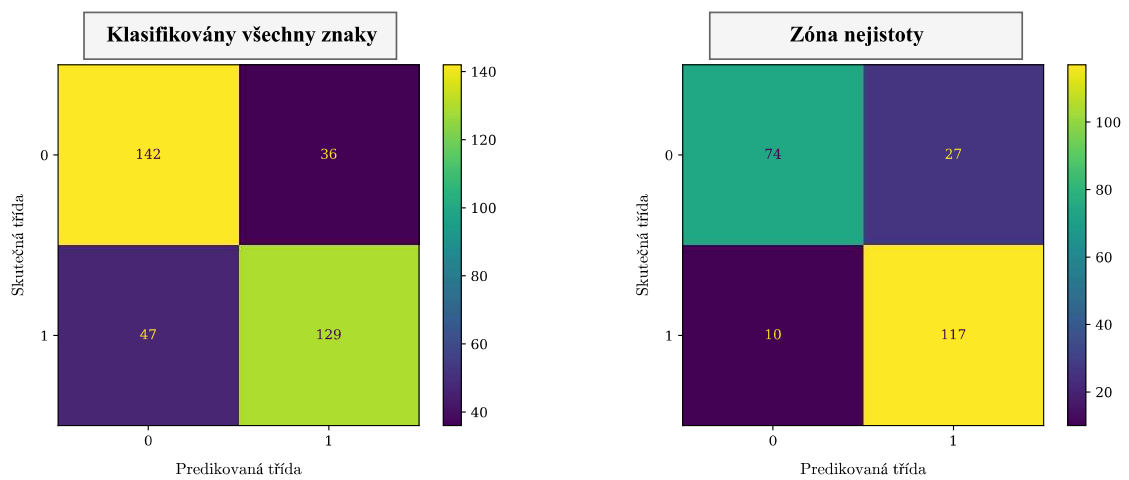
Metrika	Písmeno „k“		Písmeno „a“		Kombinace „k“ a „a“	
	Vše	Nejistota	Vše	Nejistota	Vše	Nejistota
<i>Přesnost</i>	76,1 %	89,3 %	67,2 %	67,8 %	76,6 %	83,8 %
<i>Preciznost</i>	76,4 %	87,5 %	62,8 %	65,4 %	78,1 %	81,3 %
<i>Senzitivita</i>	75,6 %	94,9 %	82,6 %	97,2 %	73,3 %	92,1 %
<i>F1 skóre</i>	76,0 %	91,1 %	71,4 %	78,2 %	75,7 %	86,3 %
<i>AUC</i>	76,1 %	88,4 %	67,4 %	60,9 %	76,5 %	82,7 %



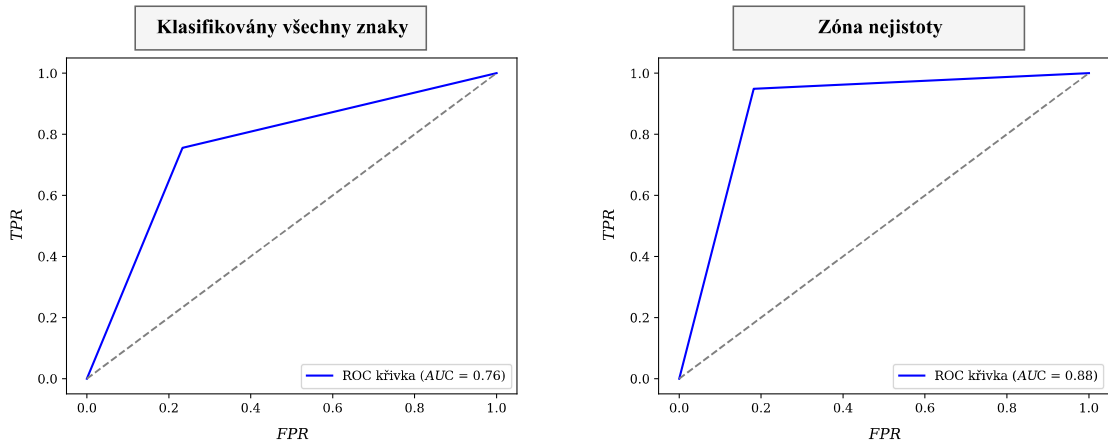
Obr. 2.10: Matice záměn CNN pro znak „k“



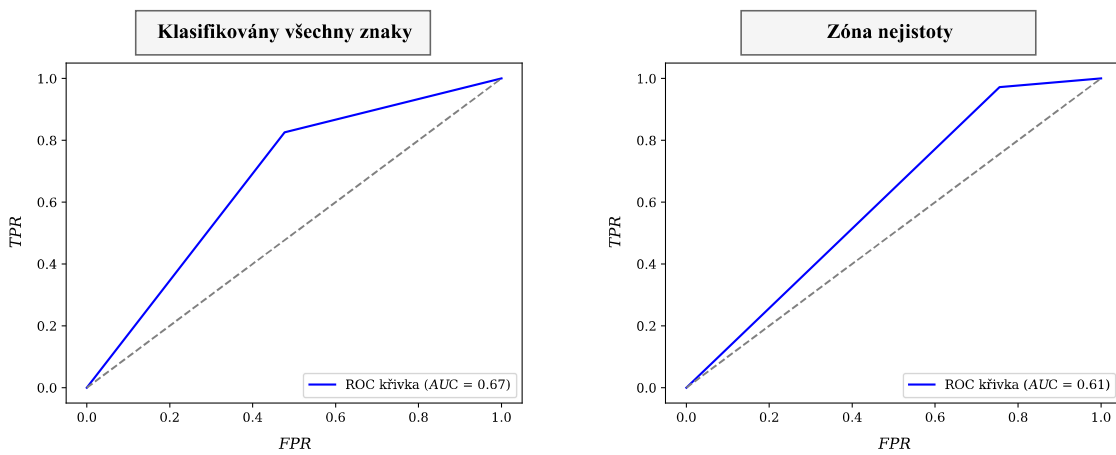
Obr. 2.11: Matice záměn CNN pro znak „a“



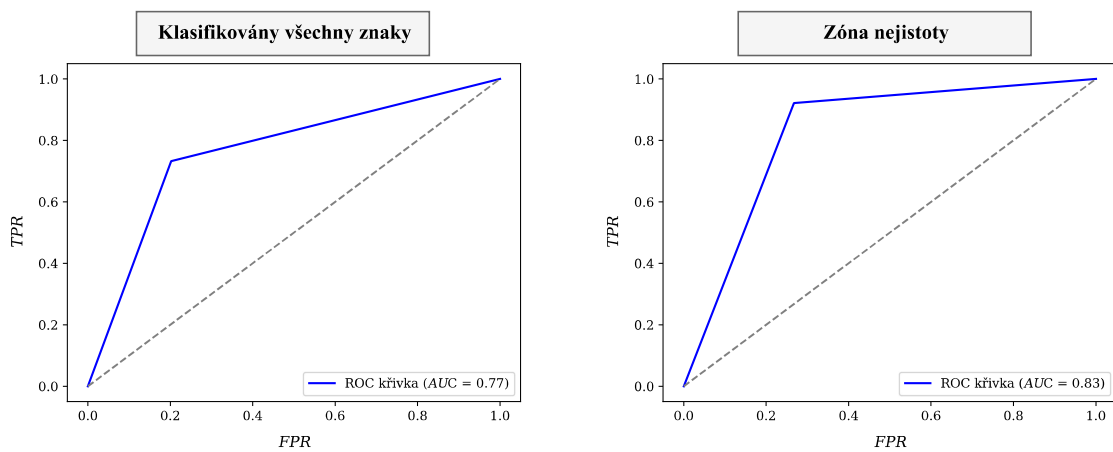
Obr. 2.12: Matice záměn CNN pro kombinaci „k“ a „a“



Obr. 2.13: ROC křivky CNN pro znak „k“



Obr. 2.14: ROC křivky CNN pro znak „a“

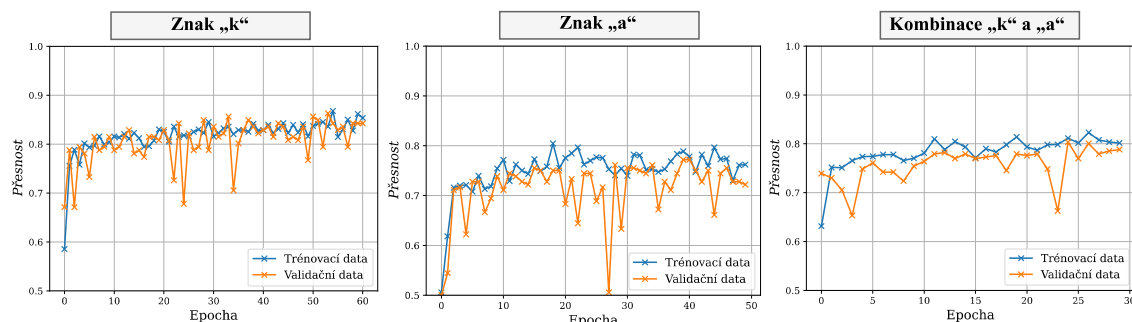


Obr. 2.15: ROC křivky CNN pro kombinaci „k“ a „a“

2.3.2 Evaluace modelu siamské sítě s autoenkodérem

Dle grafů na obrázku 2.16 je patrné, že ani v jednom případě nedochází k přetrénování modelu. Nárůst *přesnosti* zde není tak značný, jako u architektury CNN, vzhledem k využití již předtrénovaného autoenkodéru. Nejlepších výsledků při klasifikaci celé datové množiny bylo dosaženo pro písmeno „k“, a to *přesnosti* 83,3 %. Se zónou nejistoty došlo u všech datových množin k navýšení většiny metrik na přibližně 90 %, přičemž byly klasifikovány necelé 2/3 rukopisů. Modely klasifikovaly vyváženě a majoritní část hodnot se v každé matici záměn nachází na hlavní diagonálně, viz 2.17, 2.18, 2.19. Architektura byla schopna, dle 2.20, nejlépe rozlišovat mezi jednotlivými třídami pro písmeno „k“. Hodnota *AUC* činí 83,3 % při klasifikaci všech prvků a 91,4 % se zónou nejistoty, přičemž tato hodnota je obdobná i pro ostatní datové množiny, viz 2.21 a 2.22.

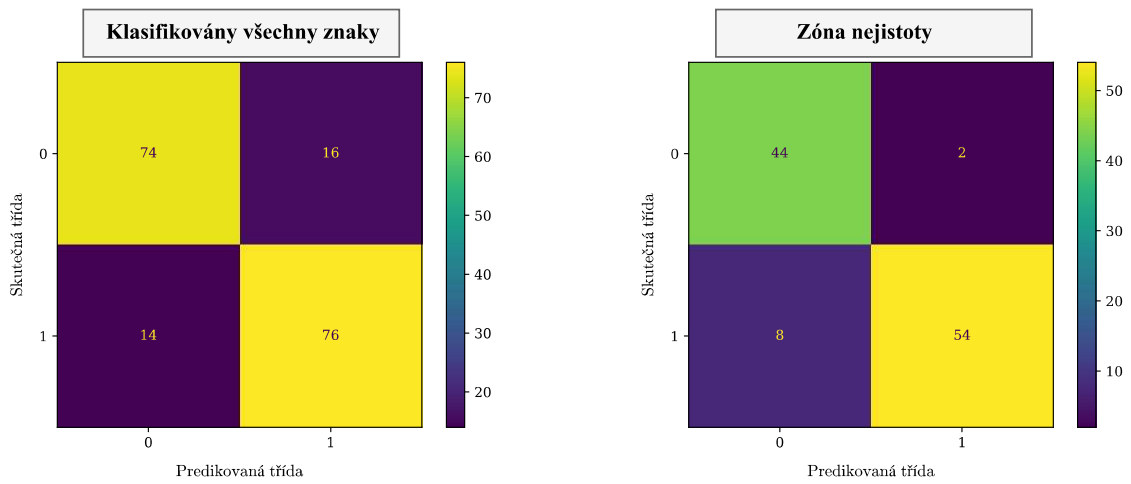
Tento model měl s využitím CNN tendenci přetrénovávat a nebyl schopen klasifikovat s *přesností* vyšší jak 50 %. Z těchto důvodů byl implementován autoenkodér. Existuje zde však určité riziko, že model se naučil verifikovat autorství čistě na základě vnější podobnosti, a ne dle specifických příznaků pisatele. Avšak výsledky, kterých bylo dosaženo v případě kombinace obou znaků tomu nenasvědčují. Implementace regularizačních technik měla negativní vliv na schopnosti modelu.



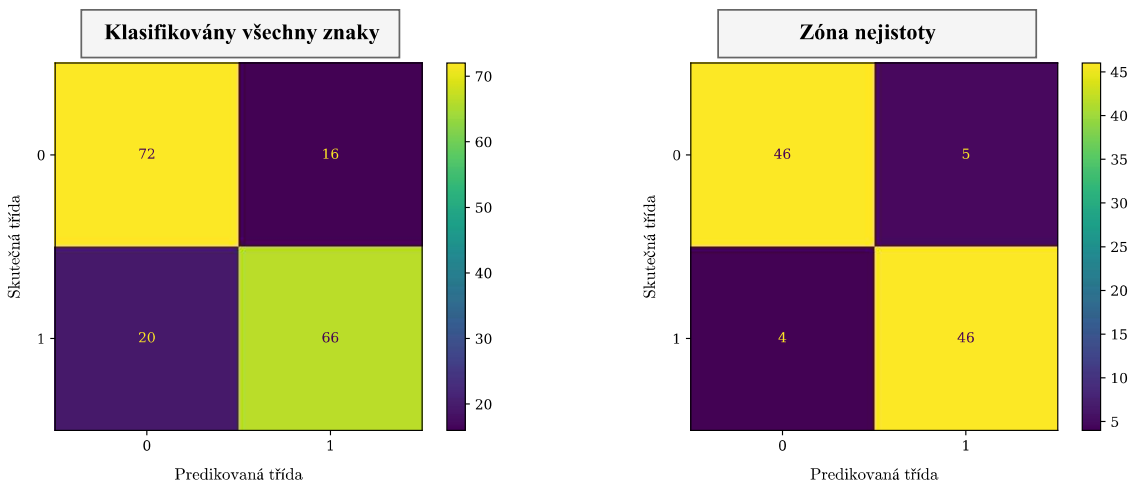
Obr. 2.16: Průběhy jednotlivých trénování siamské sítě s autoenkodérem

Tab. 2.4: Metriky siamské sítě s autoenkodérem pro písmena „k“ a „a“

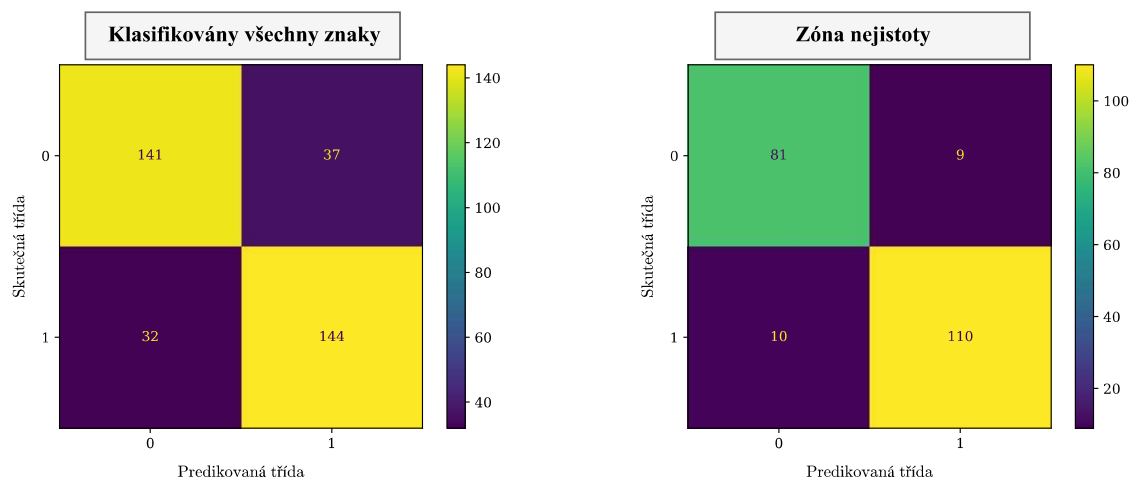
Metrika	Písmeno „k“		Písmeno „a“		Kombinace „k“ a „a“	
	Vše	Nejistota	Vše	Nejistota	Vše	Nejistota
<i>Přesnost</i>	83,3 %	90,7 %	79,3 %	91,1 %	80,5 %	91,0 %
<i>Preciznost</i>	82,6 %	96,4 %	80,5 %	90,2 %	79,6 %	92,4 %
<i>Senzitivita</i>	84,4 %	87,1 %	76,7 %	92,0 %	81,8 %	91,1 %
<i>F1 skóre</i>	83,5 %	91,5 %	78,6 %	91,7 %	80,7 %	92,0 %
<i>AUC</i>	83,3 %	91,4 %	79,3 %	91,1 %	80,5 %	90,8 %



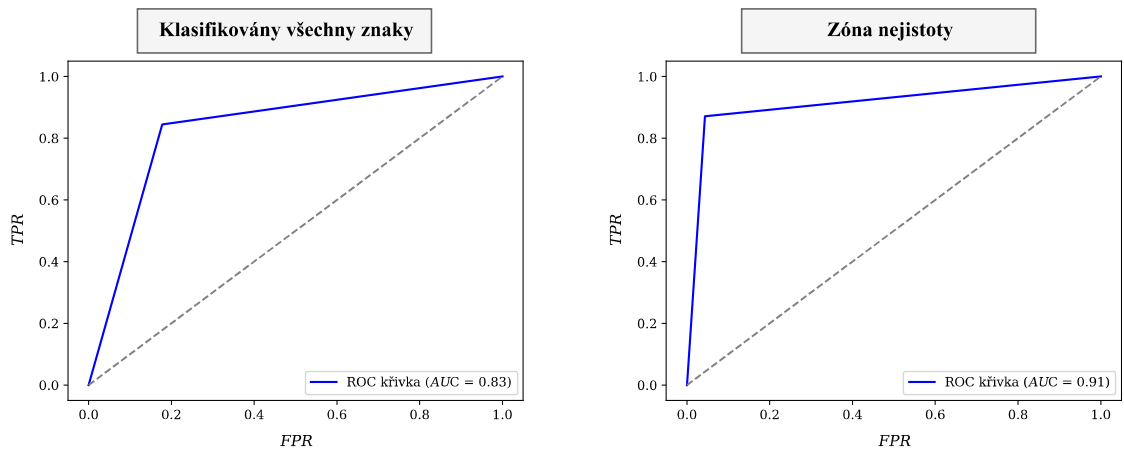
Obr. 2.17: Matice záměn siamské sítě s autoenkodérem pro znak „k“



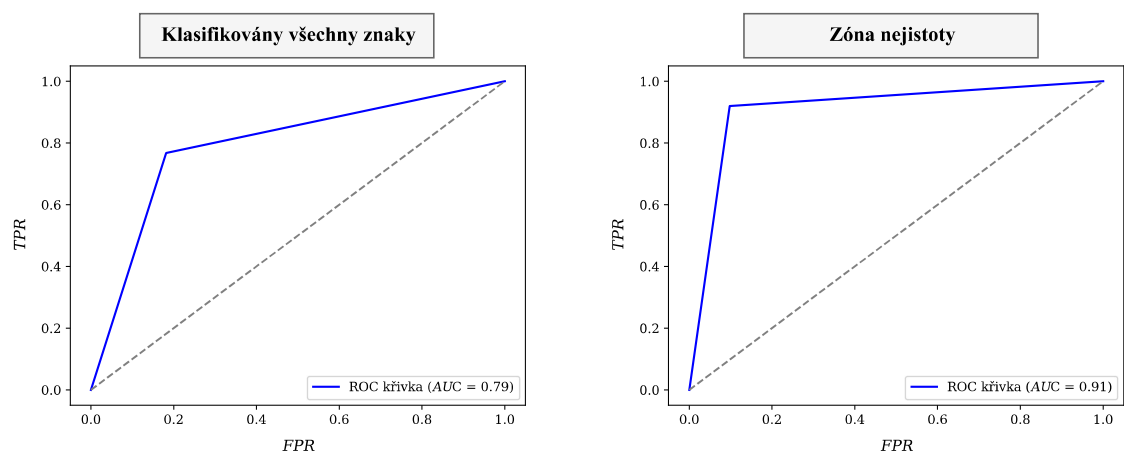
Obr. 2.18: Matice záměn siamské sítě s autoenkodérem pro znak „a“



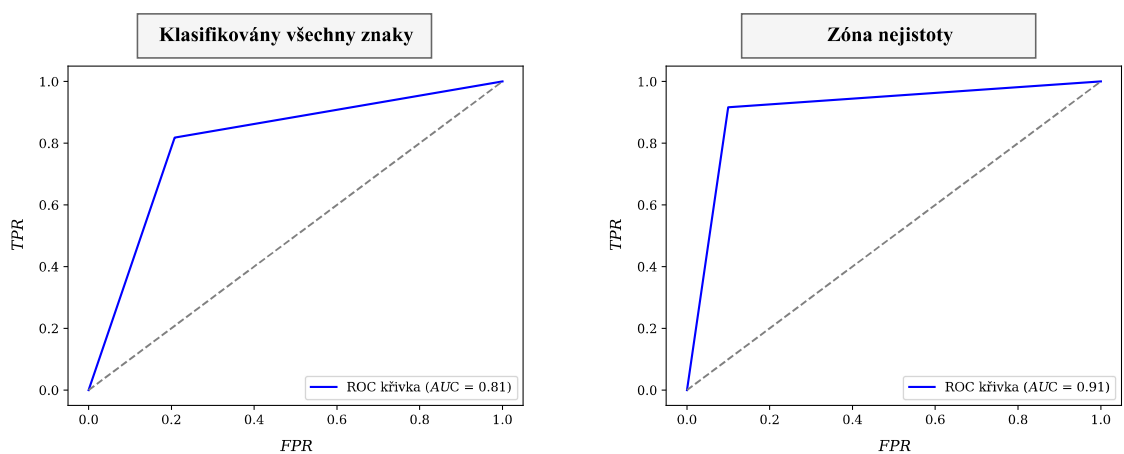
Obr. 2.19: Matice záměn siamské sítě s autoenkodérem pro kombinaci „k“ a „a“



Obr. 2.20: ROC křivky siamské sítě s autoenkodérem pro znak „k“



Obr. 2.21: ROC křivky siamské sítě s autoenkodérem pro znak „a“



Obr. 2.22: ROC křivky siamské sítě s autoenkodérem pro kombinaci „k“ a „a“

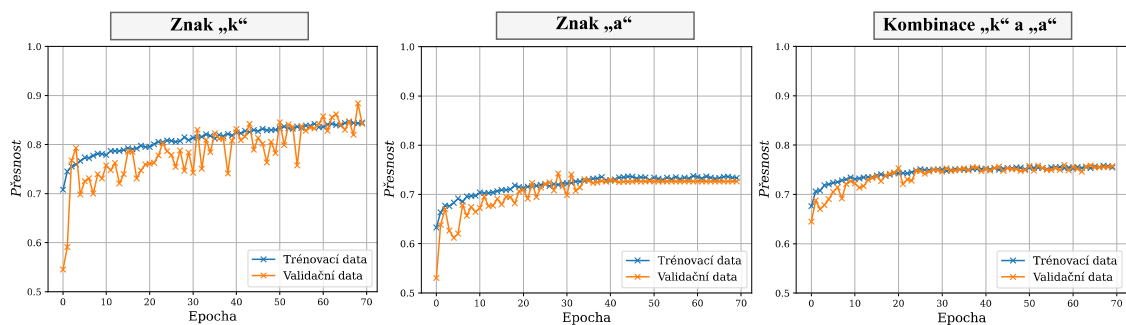
2.3.3 Evaluace modelu konvoluční siamské sítě

K největšímu nárůstu *přesnosti* došlo, dle grafů na obrázku 2.23, pro datovou množinu písmene „k“. V případě kombinace obou znaků je však tento nárůst minimální, což může být způsobeno rozsáhlejší velikostí trénovacích dávek či nižší komplexitou architektury.

Vzhledem ke skutečnosti, že i písmonožalci stanovují své závěry vždy na základě komparací většího množství vzorků, tak testování sítě bylo koncipováno následujícím způsobem. Model stanovil predikce pro jednotlivé dvojice referenčních a sporných znaků nezávisle na sobě. Ty byly posléze agregovány. Výstupem byl tedy vždy jeden štítek pro kombinaci daného počtu dvojic. Agregace predikcí byla provedena na základě zprůměrování hodnot a majoritního hlasování. Dále byl implementován přístup kombinující obě metody. Nejprve bylo provedeno majoritní hlasování. Pokud však nebyla stanovena shoda mezi pisately a průměr predikcí se blížil prahové hodnotě, pak byla stanovena nová predikce na základě tohoto průměru a nově určené prahové hodnoty. Tato rozhodovací logika byla zvolena s ohledem na tendenci modelu lépe identifikovat TN vzorky než TP vzorky. Výsledky kombinovaného hodnocení se však většinou shodovaly s těmi z majoritního hlasování, tudíž zde nejsou uvedeny. Pro lepší přehlednost se níže nachází pouze matice záměn a ROC křivky získané zprůměrováním predikcí. Vzhledem k tomu, že charakteristické vlastnosti modelu byly obdobné i při použití majoritního hlasování, jejich opětovné uvedení by bylo nadbytečné.

Z tabulek 2.5 a 2.6 vyplývá, že nejvyšší *přesnosti*, konkrétně 84,4 %, při klasifikaci všech prvků bylo dosaženo pro majoritní hlasování na datové množině znaku „k“. S využitím zprůměrování predikcí a zóny nejistoty činí *přesnost* pro tuto množinu 93,4 %, přičemž jsou správně klasifikovány více než 2/3 prvků. V případě kombinace obou množin bylo dosaženo obdobného výsledku, dále pak pro znak „a“ je *přesnost* 95,7 %. V obou případech však byly klasifikovány méně jak 2/3 prvků. Ačkoliv v případě majoritního hlasování je dosaženo nižších hodnot, byla průměrně do zóny nejistoty zařazena 1/4 prvků. Pro obě metody vyhodnocení dosahuje *senzitivita* poměrně vysokých hodnot, a to až 98,5 %. Hodnoty v maticích záměn na obrázcích 2.24, 2.25 a 2.26 inklinují k optimálnímu rozložení, čili majoritní část vzorků se nachází na hlavní diagonále. Metrika *AUC* pro písmeno „k“ nabývá 93,6 % a pro „a“ 95,9 %. Grafy ROC křivek se nachází na obrázcích 2.27, 2.28 a 2.29.

V rámci této architektury byly vyzkoušeny přístupy využívající autoenkodéru či přemostění, přičemž aktuálně právě tato konfigurace dosahuje nejlepších výsledků. Pro praktickou využitelnost by bylo vhodné model optimalizovat tak, aby došlo k redukci FP predikcí při využití zón nejistot. Teoretickým východiskem je zavedení vah při testování modelu.



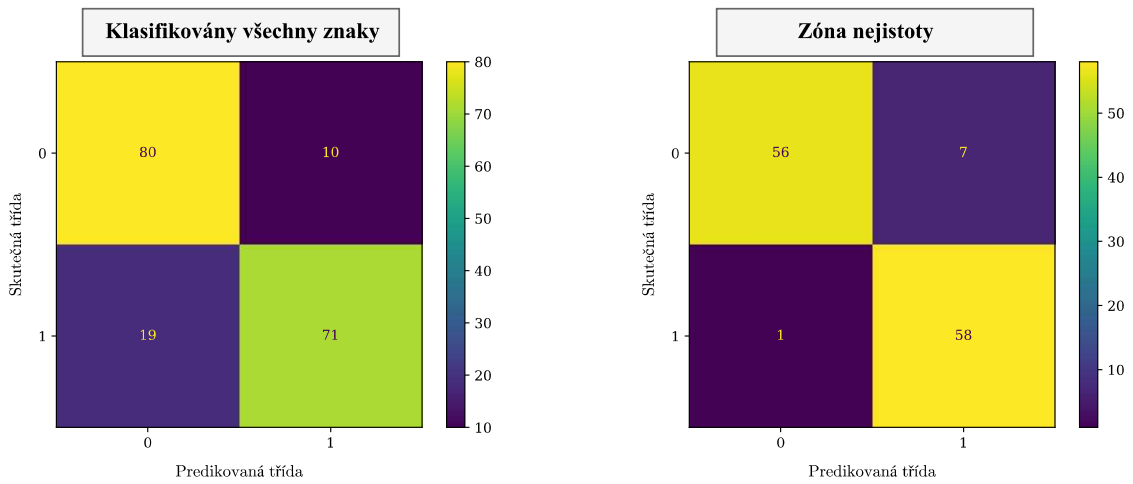
Obr. 2.23: Průběhy jednotlivých trénování konvoluční siamské sítě

Tab. 2.5: Metriky konvoluční siamské sítě pro písmena „k“ a „a“ – průměr

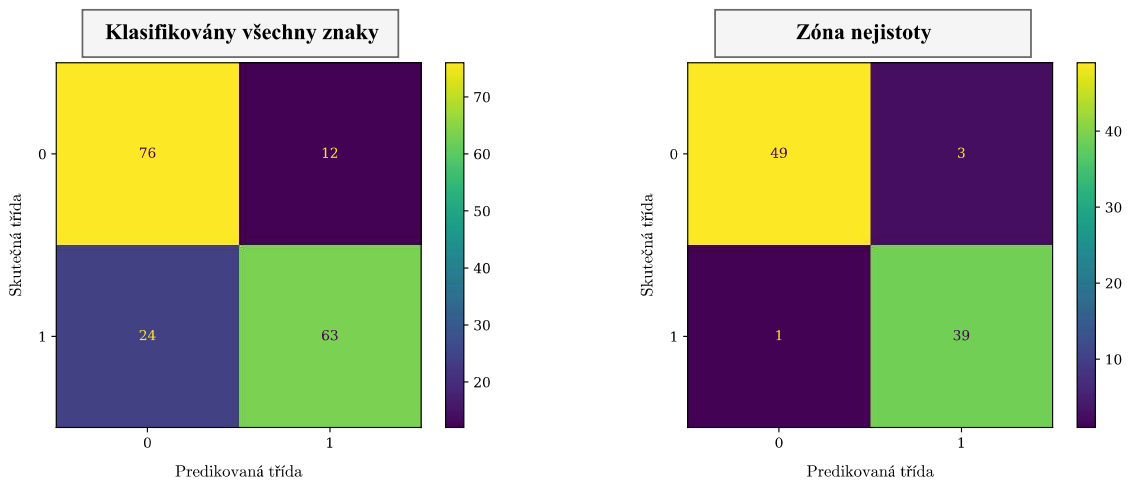
Metrika	Písmeno „k“		Písmeno „a“		Kombinace „k“ a „a“	
	Vše	Nejistota	Vše	Nejistota	Vše	Nejistota
<i>Přesnost</i>	83,9 %	93,4 %	79,4 %	95,7 %	80,9 %	93,9 %
<i>Preciznost</i>	87,7 %	89,2 %	84,0 %	92,9 %	85,2 %	90,3 %
<i>Senzitivita</i>	78,9 %	98,3 %	72,4 %	97,5 %	74,6 %	96,6 %
<i>F1 skóre</i>	83,0 %	93,6 %	77,8 %	95,1 %	79,5 %	93,3 %
<i>AUC</i>	83,9 %	93,6 %	79,4 %	95,9 %	80,8 %	94,1 %

Tab. 2.6: Metriky konvoluční siamské sítě pro písmena „k“ a „a“ – majorita

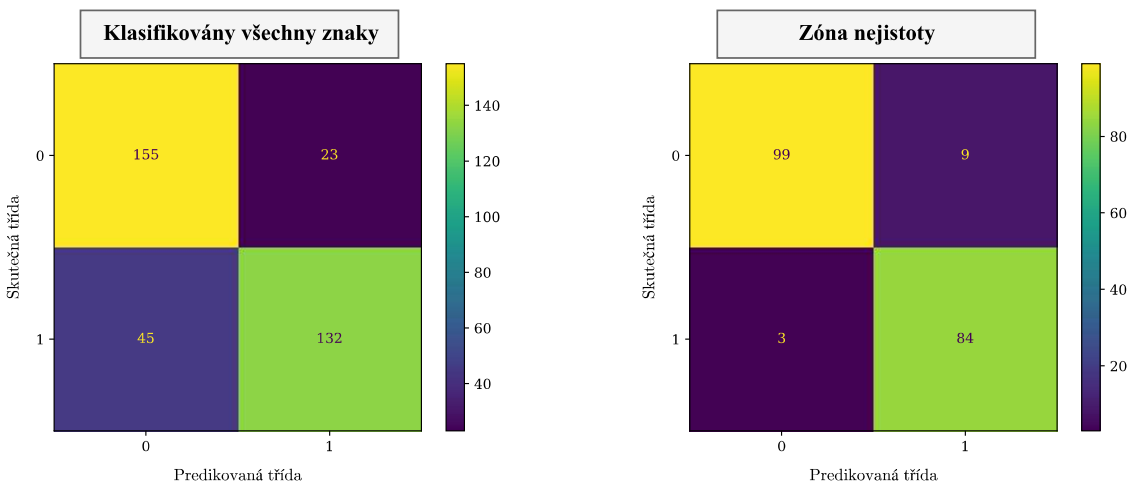
Metrika	Písmeno „k“		Písmeno „a“		Kombinace „k“ a „a“	
	Vše	Nejistota	Vše	Nejistota	Vše	Nejistota
<i>Přesnost</i>	84,4 %	89,1 %	78,9 %	88,8 %	80,3 %	89,4 %
<i>Preciznost</i>	82,3 %	85,4 %	85,7 %	83,3 %	85,4 %	87,5 %
<i>Senzitivita</i>	87,8 %	95,0 %	69,0 %	98,5 %	72,9 %	92,7 %
<i>F1 skóre</i>	85,0 %	89,9 %	76,4 %	90,3 %	78,7 %	90,0 %
<i>AUC</i>	84,4 %	89,0 %	78,8 %	88,2 %	80,3 %	89,4 %



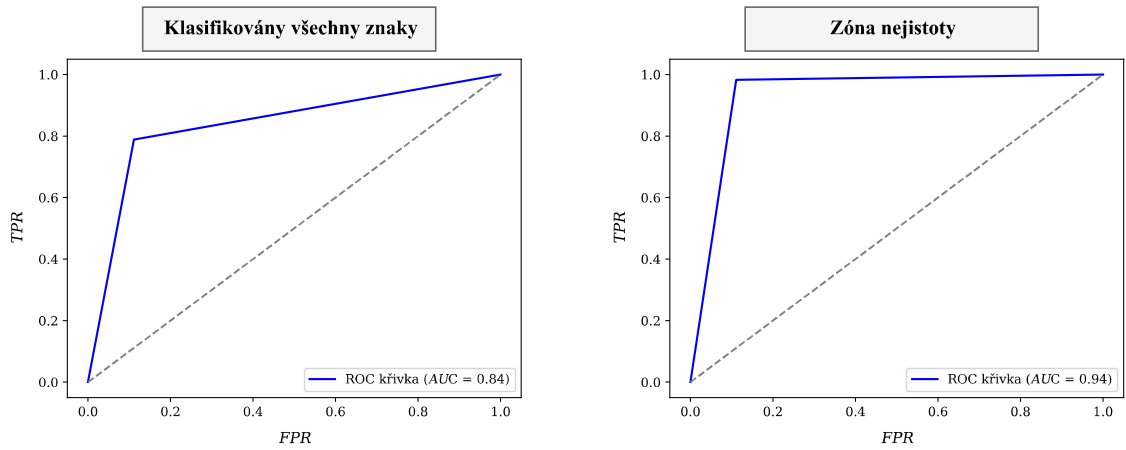
Obr. 2.24: Matice záměn konvoluční siamské sítě pro znak „k“



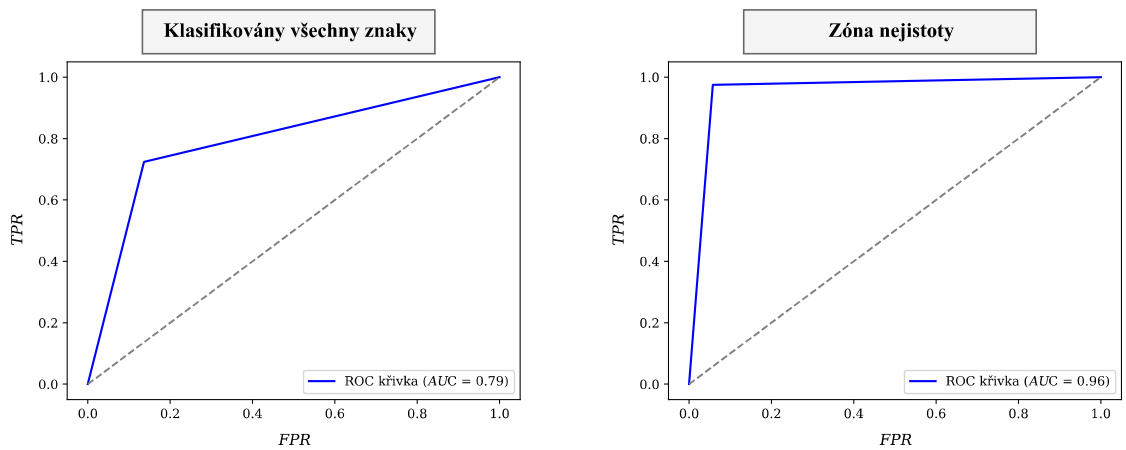
Obr. 2.25: Matice záměn konvoluční siamské sítě pro znak „a“



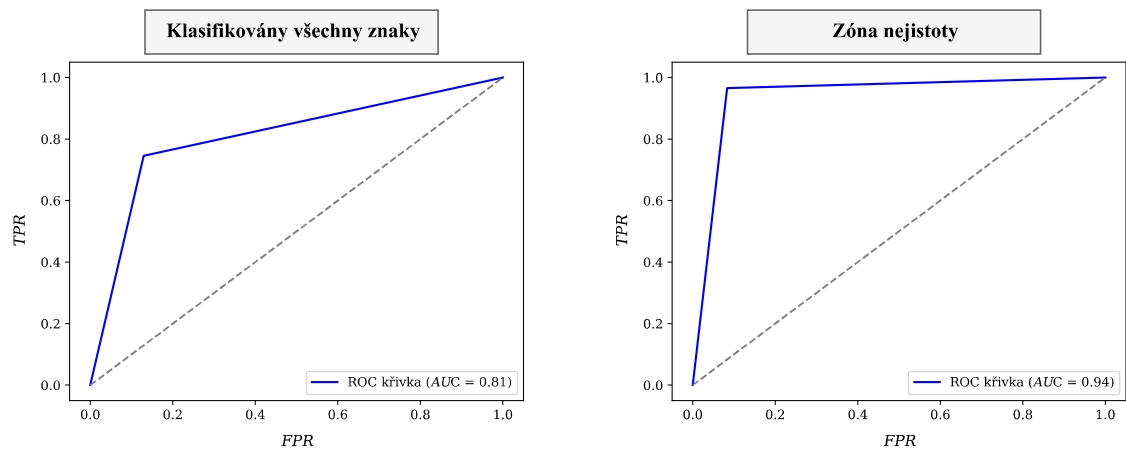
Obr. 2.26: Matice záměn konvoluční siamské sítě pro kombinaci „k“ a „a“



Obr. 2.27: ROC křivky konvoluční siamské sítě pro znak „k“



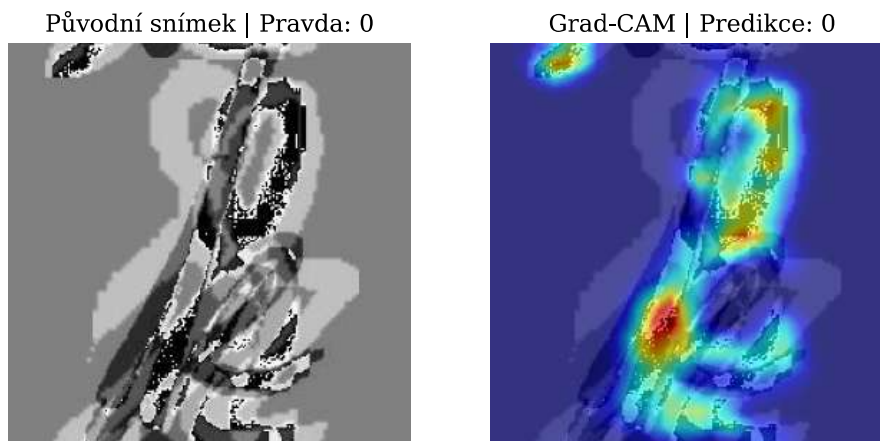
Obr. 2.28: ROC křivky konvoluční siamské sítě pro znak „a“



Obr. 2.29: ROC křivky konvoluční siamské sítě pro kombinaci „k“ a „a“

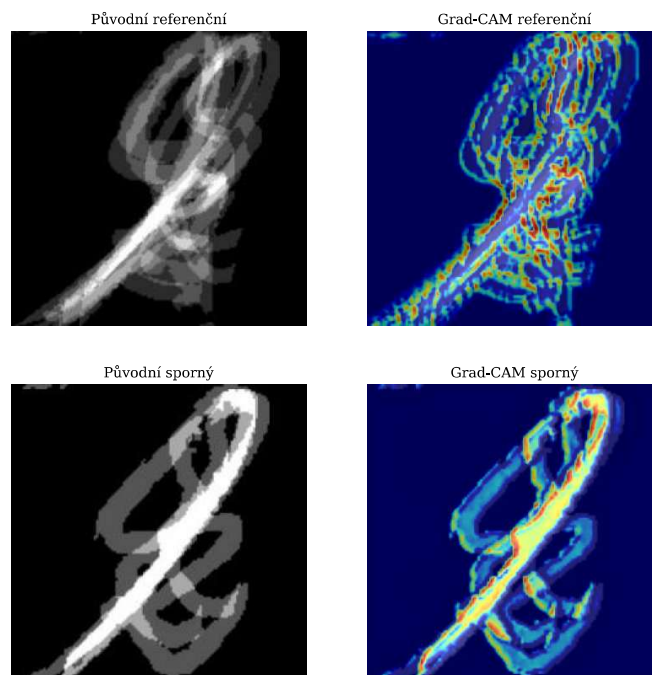
2.3.4 Výstupy teplotních map

Za účelem lepší interpretovatelnosti dosažených výsledků byly implementovány teplotní mapy Grad-CAM, a to dle principů uvedených v teoretické části v kapitole 1.6. Složitější byla situace v případě siamských sítí, kde neexistuje unifikovaný způsob, jelikož touto problematikou se příliš studií nezabývá. Výchozím zdrojem byl článek [56]. Ačkoliv teplotní mapa byla v daný okamžik generována pouze pro jeden ze vstupů siamské sítě, stejně jako u CNN, predikce pro výpočty gradientů vycházely ze zpracování dvojice snímků. Ve všech případech bylo využito barevného mapování „COLORMAP_JET“ z knihovny OpenCV, přičemž významné části snímku jsou zvýrazněny červenou barvou a naopak méně významné barvou modrou. Generované výstupy obsahují ukázkou původního snímku, teplotní mapu, skutečnou a predikovanou hodnotu štítku. I přes optimalizaci bylo poměrně náročné získat výstupy s výpovědní hodnotou, níže jsou vybrány reprezentativní ukázky. Teplotní mapy jsou uvedeny v následujícím pořadí: pro model CNN na obrázku 2.30, pro siamskou síť s autoenkodérem na obrázku 2.31 a pro konvoluční siamskou síť na obrázku 2.32. Z hlediska intuitivnosti mají největší výpovědní hodnotu výstupy získané na nezprůměrovaných dvojicích, kde je patrné, že model považoval za významné spíše jednotlivé křivky grafému než celistvé oblasti. Teplotní mapa pro CNN poukazuje na skutečnost, že architektura se zaměřovala i na rušivé elementy snímku, což poskytuje prostor pro další optimalizaci modelu, či datové množiny. Z těchto poznatků tedy plyne, že se jedná o užitečný nástroj jak pro odborníky, tak i laickou veřejnost.



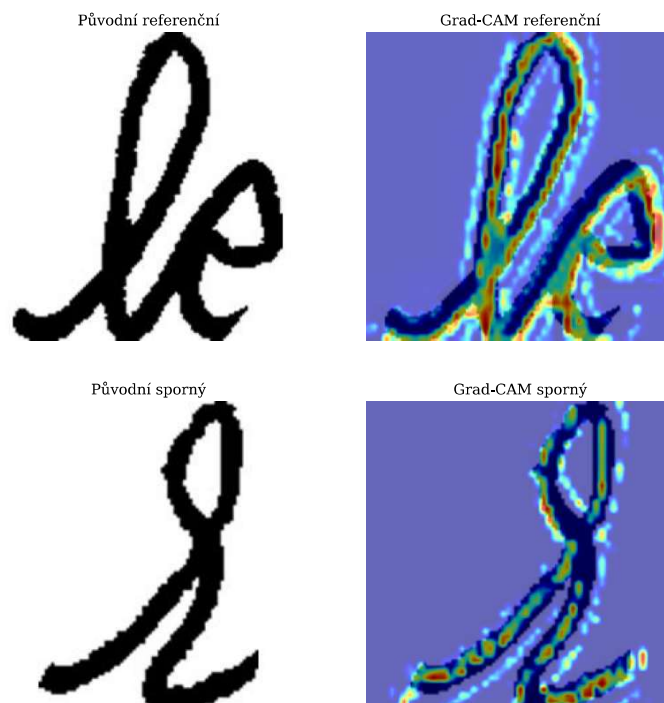
Obr. 2.30: Teplotní mapa CNN

Pravda: 1, Predikce: 1



Obr. 2.31: Teplotní mapa siamské sítě s autoenkodérem

Pravda: 0, Predikce: 0



Obr. 2.32: Teplotní mapa konvoluční siamské sítě

2.3.5 Ověření praktické využitelnosti práce

Hlavním cílem této práce bylo vytvoření podpůrného nástroje pro písmoznalce. Proto bylo nezbytné ověřit jeho praktickou využitelnost a přínosnost. Za tímto účelem byly zhotoveny dva dotazníky.

Úkolem prvního dotazníku bylo srovnat schopnost modelu verifikovat autora s laickou veřejností a písmoznalcem. Je tvořen 60 otázkami, jež obsahují variace dvojic referenčních a sporných rukopisů. Odpovědi jsou ve formátu: „Stejní autoři“, „Různí autoři“, „Nelze s jistotou určit, spíše stejní autoři“, „Nelze s jistotou určit, spíše různí autoři“. Respondenty tvoří skupina skládající se z 54 laiků a 1 písmoznalce. Zvoleným modelem je konvoluční siamská síť klasifikující dle zprůměrovaných predikcí. Tabulka 2.7 obsahuje hodnoty *přesnosti* bez využití zóny nejistoty i při jejím uplatnění, včetně počtu případů zařazených do této zóny. Pro laiky jsou uvedeny zprůměrované hodnoty spolu s příslušnými směrodatnými odchylkami. Pro tuto skupinu činí *přesnost* 83,4 %, a to při vyřazení průměrně 13 prvků. *Přesnost* modelu nabývá 92,7 %, přičemž do zóny nejistoty je zařazeno 19 vzorků. Písmoznalec dosahuje *přesnosti* 100 %, nejistých predikcí je 52. Je tedy patrné, že model prokázal schopnosti verifikovat autora lépe než laická veřejnost. S ohledem na množství vyřazených prvků modelem a písmoznalcem lze stěží srovnat dosažené *přesnosti*. Z výsledků laické veřejnosti je patrné, že jsou ve svých rozhodnutích více rezolutní, snáze tak mohou chybovat. Doba zpracování dotazníku laiky činí průměrně 18 min, u písmoznalce 22 min. Pro srovnání, model je schopen klasifikovat až 355 autorství během 15 s, případně 50 s, pokud není využit grafický procesor.

V rámci zpětné vazby písmoznalec uvádí, že je problematické analyzovat digitalizované rukopisy, které navíc kombinují různé druhy písma, mají omezené množství grafému. Nedostatečné množství rukopisu vede v praxi k vyřazení daného vzorku písmoznalcem, k čemuž došlo v rámci expertizy 1 písmoznalce v 86,7 % případů. Podstatou této práce je adaptovat se na úbytek psaného materiálu, kterému expertiza ručního písma v současnosti čelí. Model na rozdíl od respondentů při klasifikaci nedisponoval kontextovou znalostí. V každém okamžiku predikoval pouze na základě jediné dvojice, přičemž výsledná agregace probíhala čistě numerickými metodami. Naproti tomu respondenti měli v rámci každé otázky k dispozici více variant dvojic. Často se tedy stávalo, že i pro velmi odlišné rukopisy byli schopni verifikovat autorství správně, a to na základě pouze jediné výrazně podobné dvojice.

Druhý dotazník obsahuje 50 ukázek teplotních map získaných z konvoluční siamské sítě. Jeho cílem bylo potvrdit, či vyvrátit, zda existuje korelace mezi příznaky, jež písmoznalec považuje za klíčové při verifikaci autora a těmi, které jsou významné pro model. Doposud však nebyla obdržena zpětná vazba od písmoznalců, z těchto důvodů nejsou výsledky zahrnuty v této práci.

Tab. 2.7: Výsledky dotazníku srovnávajícího výstupy modelu, laiků a písmoznalce

	<i>Přesnost (vše)</i>	<i>Přesnost (nejistota)</i>	<i>Nejisté predikce</i>
Navržený model	88,3 %	92,7 %	19
Laici	80,4 % ± 7,5 %	83,4 % ± 8,1 %	13 ± 10,8 %
Písmoznalec	93,3 %	100,0 %	52

2.4 Výsledky práce v kontextu současného stavu vědy

Verifikací autora na úrovni grafému se zabývá studie [57]. Extrahování příznaků bylo provedeno prostřednictvím navržených deskriptorů. Premisou přístupu je myšlenka, že tlak vyvinutý během psaní je proměnlivý. Přičemž tato proměnlivost se opakuje na stejných místech různých vzorků grafému pro daného pisatele. Tyto vzory se nejvíce projevují v šířce tahu a různých stupních šedi. Jednotlivé deskriptory byly tedy navrženy tak, aby se zaměřovaly například na stupně šedi v kostře grafému, průměrnou hodnotu stupňů šedi v kolmém směru ke kostře grafému, či šířku tahu. Bylo zde dosaženo *přesnosti* až 98,4 %, a to pro znak „U“. Datová množina vznikla za kontrolovaných podmínek. Obsahuje rukopisy 50 osob, kde každá poskytla 50 vzorků znaků: „C“, „S“, „∩“, „U“, „~“. Autoři vepisovali jednotlivé znaky do mřížky o velikosti buňky 2 × 2 cm, a to jednotným psacím nástrojem. Rozlišení použitého skeneru bylo 1 200 dpi a výsledné snímky nabývají rozměru 800 × 800 pixelů. V rámci předzpracování bylo aplikováno Otsuovo prahování a eroze. Vzhledem k okolnostem vzniku datové množiny je praktická využitelnost tohoto přístupu diskutabilní. Autentické rukopisy se mnohdy nachází na různých podkladových materiálech a nabývají rozdílných kvalit, čímž jsou značně zvýšeny nároky na kvalitu předzpracování dat. V praxi je tedy poměrně obtížné získat natolik kvalitní vzorky, aby mohlo být dosaženo takové úrovně rozlišení.

Studie [47] srovnává různé přístupy v oblasti expertizy ručního písma, které se zaměřují primárně na komparaci rukopisů. V rámci práce jsou využity celkem tři extraktory příznaků: ViT 16patch, EfficientNet B0 a ResNet 18. Ty jsou implementovány v rámci autoenkodéru a siamské sítě. Cílem obou architektur je získání latentních prostorů, jež jsou následně komparovány prostřednictvím kosinové podobnosti. Rukopisy byly opatřeny z veřejných datových množin CVL [58] a CSAFE [59], které obsahují texty v anglickém a německém jazyce. Jednotlivé sporné a referenční dokumenty obsahují různé délky textu v rozsahu od 14 do 86 slov. Ze studie vyplývá, že nejlepších výsledků bylo dosaženo při využití autoenkodéru v kombinaci s architekturou ViT. Pro datovou množinu CSAFE činí nejvyšší dosažená hodnota *AUC* 98,0 % a *přesnost* 94,45 %. Hodnota *AUC* pro datovou množinu CVL [58] nabývá 99,0 %, přičemž *přesnost* je 97,15 %. Ve studii je rovněž provedeno srovnání s ostatními

přístupy, konkrétně s metodami využívajícími shlukování a hierarchický bayesovský model. V práci je uvedeno, že metoda založená na hlubokém učení je robustnější vůči kratší délce textu. Přestože zde bylo dosaženo vysoké *přesnosti*, délky použitých textů neodpovídají současným potřebám expertizy ručního písma. Současně však tyto poznatky potvrzují smyslupnost využití hlubokého učení v rámci této disciplíny.

Analýze ručně psaných podpisů se věnuje zdroj [48]. Architektura zde použita je tvořena dvěma moduly, přičemž koncept vychází z mechanismů siamské sítě. Páteřní část architektury se skládá ze 4 vrstev obsahujících vždy 2 konvoluční vrstvy, normalizaci dávek a podvzorkovací vrstvu typu Generalized Mean. První modul je určen k verifikaci autora. Za tímto účelem je zde využito křížové pozornosti (cross-attention), jejíž předností je schopnost vytvořit kontext mezi srovnávanými podpisy. Snímky jsou nejprve zpracovány páteřní sítí, následně je aplikována křížová pozornost, a takto získané vektory příznaků jsou v rámci klasifikace zpracovány binární křížovou entropií. Druhý modul vznikl s ohledem na skutečnost, že klasifikace pouze na základě nejvýznamějších příznaků nemusí dostatečně pokrývat variabilitu pisatele. Bylo tedy využito bilineárních operací, jež umožnily stanovit komplementárnost mezi různými částmi map příznaků. Zde byly vstupem pouze originální podpisy. Trénování probíhalo s využitím obou modulů. Pro účely testování bylo využito modulu zaměřeného na verifikaci. Vstupní snímky měly rozměr 155×220 pixelů. Bylo dosaženo *přesnosti* 100,0 % pro datovou množinu CEDAR [15] obsahující podpisy od 55 autorů, přičemž vznikla za kontrolovaných podmínek. Na datové množině BHSig-B [60] tvořené 100 pisateli *přesnost* činí 69,0 %. Vzhledem k délce rukopisu, na kterém byla analýza provedena, nelze srovnat dosažené výsledky s touto prací. Současně se však jedná o poměrně inovativní přístup, který cílí na zachování co největšího množství extrahovaných příznaků.

Existuje četné množství studií, jež se zabývají využitím hlubokého učení v expertize ručního písma, avšak většina z nich je limitována faktem, že modely jsou navrženy primárně pro analýzu rukopisů na úrovni vět či slov. Jedinečnost této práce je dána skutečností, že referenční a sporný materiál je tvořen pouze variacemi jediného písmene „k“, případně „a“. Znak „k“ je záměrně zvolen s ohledem na jeho komplexní strukturu, čímž ho lze považovat za nositele významných informací o pisateli. Tento přístup je koncipován tak, aby co nejlépe refletoval aktuální požadavky expertizy ručního písma a byl schopen poskytovat uspokojivé výsledky i při nedostatku grafémů. Je nezbytné zmínit, že modely jsou trénovány na rukopisech českých a slovenských autorů, které vykazují jistá specifika oproti cizojazyčným datovým množinám. Klíčové je využití autentické množiny, jež obsahuje jak tiskací, tak psací znaky. Ačkoliv studie [57] analyzuje rukopisy na úrovni grafému, jejím hlavním nedostatek je datová množina vytvořená za kontrolovaných podmínek.

Závěr

Hlavním cílem bakalářské práce byl návrh a implementace tří architektur neuronových sítí, jakožto podpůrných nástrojů pro účely expertizy ručního písma. Současně byla provedena analýza nynějšího stavu této disciplíny, jež odhalila stěžejní problémy, kterými je subjektivní povaha hodnocení, absence kvantitativních technik, statisticky obtížně zpracovatelné výstupy a úbytek grafémů. Na základě rešerše byly identifikovány vhodné moderní metody založené na hlubokém učení.

Cíle práce byly splněny a současně bylo prokázáno, že expertiza ručního písma může být realizována nejen na úrovni větných celků či slov, ale i na úrovni grafému. Navržené modely využívají odlišné přístupy předzpracování vstupních dat, přičemž nejlepší výsledky poskytla metoda založená na komparaci nezprůměrovaných referenčních a sporných písmen, která byla implementována v rámci konvoluční siamské sítě. S využitím zón nejistot bylo dosaženo *přesnosti* 93,4 % pro znak „k“, 95,7 % v případě znaku „a“ a 93,9 % pro kombinaci obou znaků. Hodnota *AUC* ve všech případech činí přibližně 94,0 %. Tyto výsledky byly získány na autentických testovacích množinách obsahujících 180 vzorků „k“ a 175 „a“ od celkem 45 pisatelů. Vzhledem ke skutečnosti, že nedošlo k zásadnímu poklesu metrik v případě sloučení datových množin obou písmen, lze tuto architekturu považovat za potenciálně robustní vůči různým znakům. Tato skutečnost je indikátorem, že se model nezaměřuje pouze na vnější podobnost grafému, ale i na jeho interní příznaky. Za účelem lepší interpretovatelnosti dosažených výsledků byly vytvořeny teplotní mapy Grad-CAM. Tato práce byla prezentována na studentské konferenci Student EEICT, kde se umístila na prvním místě.

Hlavním přínosem této bakalářské práce je schopnost verifikovat autora i při omezeném množství autentických rukopisů. Navržená konvoluční siamská síť vykazuje vyšší *přesnost* než laická veřejnost, a to o 9,3 %. Současně poskytuje více jistých predikcí ve srovnání s certifikovaným písmoznalcem, ve více než polovině případů a v mnohonásobně kratším čase. Jedná se o jedinečný výstup, jelikož studie odpovídající současnému stavu vědy jsou převážně orientovány na zpracování většího množství textu či analyzují laboratorně vytvořené rukopisy, v důsledku čehož je limitována jejich praktická využitelnost.

Výhodou tohoto přístupu je schopnost modelu automaticky detekovat významné příznaky písma a provést následnou klasifikaci v rámci jedné ucelené architektury. Dosažené výsledky jsou objektivní, reprodukovatelné a jednoduše statisticky zpracovatelné, což přispívá k vyšší míře konzistentnosti důkazního materiálu. Schopnost stanovit autorství pouze na základě variací jediného znaku je vzhledem k úbytku psaného projevu klíčová. Možnost kvantitativního zpracování dat představuje časovou úsporu pro experty při ověřování jimi dosažených kategorických závěrů.

Současně však nelze opomíjet nevýhody s tímto přístupem spojené. Vlivem digitalizace a následného předzpracování dat dochází ke ztrátě dynamických informací o písmu, které jsou pro písmoznalce velmi cenné.

V rámci dalšího rozvoje práce by bylo užitečné rozšířit datové množiny o další znaky. Dále by bylo vhodné vyzkoušet souborové učení (ensemble learning). Tento přístup může být přínosný vzhledem k rozdílnosti metod předzpracování snímků pro jednotlivé modely.

Literatura

- [1] SAPNA, Aggarwal. "Handwriting & Individuality": A Psychological, Biomechanical and Growth & Developmental Perspective of Handwriting. *The International Journal of Indian Psychology*. 2015, vol. 3, no. 8, s. 1–4. ISSN 2349-3429.
- [2] FERNANDES, Carina Pereira; MONTALVO, Gemma; CALIGIURI, Michael; PERTSINAKIS, Michael; GUIMARÃES, Joana. Handwriting Changes in Alzheimer's Disease: A Systematic Review. *Journal of Alzheimer's Disease*. 2023, vol. 96, no. 1, s. 1–11. ISSN 1875-8908.
- [3] KOWAL, Durlabh Singh; GUPTA, Pradip Kumar. Handwriting analysis: A psychopathic viewpoint. *The International Journal of Indian Psychology*. 2021, vol. 9, no. 1, s. 1–7. ISSN 2349-3429.
- [4] HICKLIN, R. Austin; EISENHART, Linda; RICHETELLI, Nicole; MILLER, Meredith D.; BELCASTRO, Peter et al. Accuracy and Reliability of Forensic Handwriting Comparisons. *Proceedings of the National Academy of Sciences*. 2022, vol. 119, article e2119944119. ISSN 1091-6490.
- [5] SUKHIJA, Ranveer; KUMAR, Munish; JINDAL, M. K. Document forgery detection: A comprehensive review. *International Journal of Data Science and Analytics*. 2025, s. 1–23. ISSN 2364-4168.
- [6] TAYLOR, Melissa; BIRD, Carolyne; BISHOP, Brett; BURKES, Ted; CALIGIURI, Michael P. et al. *Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach*. Online. NIST. Dostupné z: <https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8282.pdf>. [cit. 2024-11-18]
- [7] BREYER Stephen; BERGER, Margaret A.; GOODSTEIN, David; GIANNELLI, Paul C.; IMWINKELRIED, Edward J. et al. *Reference Manual on Scientific Evidence*. Third Edition. Washington, DC: The National Academies Press, 2011. ISBN 978-0-309-21421-6.
- [8] DARGAN Shaveta; KUMAR Munish. Writer Identification System Based on Offline Handwritten Text in Gurumukhi Script. In: *2020 Sixth International Conference on Parallel, Distributed and Grid Computing*. 2020, s. 544–549. ISBN 978-1-7281-7132-6.

- [9] ALAEI, Fahimeh a ALAEI, Alireza. Handwriting analysis: Applications in person identification and forensic. In: DAIMI, Kevin, FRANCA III, Guillermo a ENCINAS, Luis Hernández, ed. *Breakthroughs in Digital Biometrics and Forensics*. Cham: Springer International Publishing, 2022, s. 147–165. ISBN 978-3-031-10706-1.
- [10] MORRIS, Ron; *Forensic Handwriting Identification: Fundamental Concepts and Principles*. Second Edition. Cambridge, Massachusetts: Academic Press, 2020. ISBN 978-0-12-409602-8.
- [11] XIAO, Wanghui. Offline Handwritten Signature Verification using Siamese Network Model. In: *2024 6th International Conference on Electronic Engineering and Informatics (EEI)*. 2024, s. 1536–1539. ISBN 979-8-3503-5359-4.
- [12] AGIUS, Anna; JONES, Kylie; EPPLE, Rochelle; MORELATO, Marie; MORET, Sébastien et al. The Use of Handwriting Examinations beyond the Traditional Court Purpose. *Science & Justice*. 2017, vol. 57, no. 5, s. 394–400. ISSN 1355-0306.
- [13] YANG, Jing; SHOKOUHIFAR, Mohammad; YEE, Por Lip; KHAN, Abdullah Ayub; AWAIS, Muhammad et al. DT2F-TLNet: A novel text-independent writer identification and verification model using a combination of deep type-2 fuzzy architecture and Transfer Learning networks based on handwriting data. *Expert Systems with Applications*. 2024, vol. 242, article 122704. ISSN 0957-4174.
- [14] KAUR, Harmandeep; KUMAR, Munish. Signature identification and verification techniques: state-of-the-art work. *Journal of Ambient Intelligence and Humanized Computing*. 2023, vol. 14, no. 2, s. 1027–1045. ISSN 1868-5145.
- [15] PRAKASH, Shreelakshmi G. *CEDAR-dataset*. Online. Kaggle. Dostupné z: <https://www.kaggle.com/datasets/shreelakshmigp/cedardataset>. [cit. 2025-05-09].
- [16] BIDA LAB. *MCYT-SignatureOff-75*. Online. BiDA Lab – Biometrics and Data Pattern Analytics Research Group. Dostupné z https://bidalab.eps.uam.es/listdatabases?id=MCYT_SignatureOff_75#page. [cit. 2025-05-09].
- [17] TREHAN, K. a RANI, R. Review on Writer Identification and Verification Methods. In: *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. 2023, s. 1–6. ISBN 979-8-3503-3509-5.

- [18] AKINROLABU, Olatunde David; AJAYI, Olusola Olajide; EBITIGHA, Akinola Elijah; ADEGBITE, Adewuyi Adetayo; OBAFEMI, Joy Rotimi et al. Handwritten Signature Verification Model Using Transfer Deep Learning Technique. In: *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*. 2024, s. 1–6. ISBN 979-8-3503-5815-5.
- [19] INUNGANBI, Sanasam. A systematic review on handwritten document analysis and recognition. *Multimedia Tools and Applications*. 2024, vol. 83, no. 2, s. 5387–5413. ISSN 1380-7501.
- [20] YING, Xue. An overview of overfitting and its solutions. *Journal of physics: Conference series*. 2019, vol. 1168, no. 2, article 022022. ISSN 1742-6596.
- [21] GONZALEZ, Rafael C., WOODS, Richard E. *Digital Image Processing*. Fourth Edition. New York, NY: Pearson, 2017. ISBN 978-0-13-335672-4.
- [22] HAGSTRÖM Adrian Leo; STANIKZAI Rustam; BIGUN Josef; ALONSO-FERNANDEZ Fernando. Writer Recognition Using Off-Line Handwritten Single Block Characters. In: *2022 International Workshop on Biometrics and Forensics (IWBF)*. 2022, s. 1–6. ISBN 978-1-6654-6962-3.
- [23] PRASAD, Preena; ANITHA, J. a ANIL, Divapriya. A Systematic Review of Noise Types, Denoising Methods, and Evaluation Metrics in Images. In: *2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*. 2023, s. 1–9. ISBN: 979-8-3503-4167-6.
- [24] SIMARD, P. Y.; STEINKRAUS, D.; PLATT, J. C. Best practices for convolutional neural networks applied to visual document analysis. In: *Seventh International Conference on Document Analysis and Recognition*. 2003, s. 958–963. ISBN 0-7695-1960-1.
- [25] NGUYEN, Pham Sy; HUY, Tran Nhat; TUAN, Tong Anh; TRUNG, Pham Duy a LONG, Hoang Viet. Hybrid feature extraction and integrated deep learning for cloud-based malware detection. *Computers & Security*. 2025, vol. 150, article 104233. ISSN 0167-4048.
- [26] RAINIO, Oona; TEUHO, Jarmo a KLÉN, Riku. Evaluation metrics and statistical tests for machine learning. *Scientific Reports*. 2024, vol. 14, article 6086. ISSN 2045-2322.

- [27] CANBEK, Gürol; TEMIZEL, Tugba Taskaya; SAGIROGLU, Seref. BenchMetrics: A systematic benchmarking method for binary classification performance metrics. *Neural Computing and Applications*. 2021, vol. 33, no. 21, s. 14623–14650. ISSN 1433-3058.
- [28] LOWE, David G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*. 2004, vol. 60, s. 91–110. ISSN 1573-1405.
- [29] OJALA, T.; PIETIKAINEN, M. a MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002, vol. 24, no. 7, s. 971–987. ISSN: 1939-3539.
- [30] WAN, Qian; ZOU, Qin. Learning Metric Features for Writer-Independent Signature Verification using Dual Triplet Loss. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021, s. 3853–3859. ISBN 978-1-7281-8808-9.
- [31] PRINCE, Simon J.D. *Understanding Deep Learning*. Cambridge, Massachusetts: The MIT Press, 2023. ISBN 9780262377102.
- [32] SUZUKI, Kenji; *Artificial Neural Networks - Methodological Advances and Biomedical Applications*. Rijeka: InTechOpen, 2011. ISBN 978-953-307-243-2.
- [33] GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. Cambridge, Massachusetts: The MIT Press, 2016. ISBN 0262035618.
- [34] ZHAO, Xia; WANG, Limin; ZHANG, Yufei; HAN, Xuming; DEVECI, Muhammet et al. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*. 2024, vol. 57, no. 4, article 99. ISSN 1573-7462.
- [35] HUBEL, David H.; WIESEL, Torsten N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*. 1968, vol. 195, no. 1, s. 215–243. ISSN 1469-7793.
- [36] TAKAHASHI, Satoshi; SAKAGUCHI, Yusuke; KOUNO, Nobuji; TAKASAWA, Ken; ISHIZU, Kenichi et al. Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. *Journal of Medical Systems*. 2024, vol. 48, no. 1, s. 1–22. ISSN 1573-689X.

- [37] YANG, Wenchuan; YU, Haoran; CUI, Baojiang; SUI, Runqi a GU, Tianyu. Deep neural network pruning method based on sensitive layers and reinforcement learning. *Artificial Intelligence Review*. 2023, vol. 56, no. 2, s. 1897–1917. ISSN 1573-7462.
- [38] CAO, Qizhi; ZHANG, Kaibing; XUE, Dinghua a ZHANG, Zhouqiang. Focusing on Significant Guidance: Preliminary Knowledge Guided Distillation. In: LIN, Zhouchen et al., ed. *Pattern Recognition and Computer Vision: PRCV 2024. Lecture Notes in Computer Science*, vol. 15033. Singapore: Springer, 2025, s. 304–318. ISBN 978-981-97-8502-5.
- [39] SUDHARSHAN, Duth P. a VISMAYA, R. N. Handwritten Signature Verification System using Deep Learning. In: *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*. 2022, s. 1–5. ISBN: 978-1-6654-9801-2.
- [40] BROMLEY, Jane; GUYON, Isabelle; LECUN, Yann; SÄCKINGER, Eduard; SHAH, Roopak. Signature verification using a "siamese" time delay neural network. In: *Advances in Neural Information Processing Systems 6 (7th NIPS Conference)*. 1993, s. 737–744. ISBN 1-55860-322-0.
- [41] VALERO-MAS, Jose J.; GALLEGO, Antonio Javier; RICO-JUAN, Juan Ramón. An overview of ensemble and feature learning in few-shot image classification using siamese networks. *Multimedia Tools and Applications*. 2024, vol. 83, no. 7, s. 19929–19952. ISSN 1380-7501.
- [42] GU, Jiuxiang; WANG, Zhenhua; KUEN, Jason; MA, Lianyang; SHAHROUDY, Amir et al. Recent advances in convolutional neural networks. *Pattern Recognition*. 2018, vol. 77, s. 354–377. ISSN 0031-3203.
- [43] DLAMINI, Nkosikhona; ZYL, Terence L van. Author Identification from Handwritten Characters using Siamese CNN. In: *2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*. 2019, s. 1–6. ISBN 978-1-7281-0040-1.
- [44] DOSOVITSKIY, Alexey; BEYER, Lucas; KOLESNIKOV, Alexander; WEISSENBERN, Dirk; ZHAI, Xiaohua et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Online. In: *International Conference on Learning Representations*. 2021, s. 1–21. Dostupné z: <https://openreview.net/pdf?id=YicbFdNTTy>. [cit. 2024-12-7]

- [45] VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion et al. Attention is All You Need. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*. 2017, s. 1–11. ISBN 978-1-5108-6096-4.
- [46] KOEPF, Michael; KLEBER, Florian; SABLATNIG, Robert. Writer Identification and Writer Retrieval Using Vision Transformer for Forensic Documents. In: *Document Analysis Systems, 15th IAPR International Workshop*. 2022, s. 352–366. ISBN 978-3-031-06555-2.
- [47] KIM, Juhyeon; PARK, Soyoung; CARRIQUIRY, Alicia. A deep learning approach for the comparison of handwritten documents using latent feature vectors. *Statistical Analysis and Data Mining: The ASA Data Science Journal*. 2024, vol. 17, no. 1, article e11660. ISSN 1932-1872.
- [48] MENG, Haotian; LIN, Xiaoya; UBUL, Kurban a AYSA, Alimjan. Multi-Task Interaction Network Based on a Cross-Attention Fusion Mechanism for Offline Signature Verification. In: *International Conference on Pattern Recognition*. 2025, s. 17–31. ISBN 978-3-031-78389-0.
- [49] SIMONYAN, Karen a ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. Online, preprint. In: *International Conference on Learning Representations (ICLR 2015)*. Dostupné z: <https://arxiv.org/pdf/1409.1556>. [cit. 2025-05-09].
- [50] HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing a SUN, Jian. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, s. 770–778. ISBN: 978-1-4673-8851-1.
- [51] TAN, Mingxing a LE, Quoc. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*. 2019, vol. 97, s. 6105–6114. Dostupné z: <https://proceedings.mlr.press/v97/tan19a.html> [cit. 2025-01-28].
- [52] SELVARAJU, Ramprasaath R.; COGSWELL, Michael; DAS, Abhishek; VEDANTAM, Ramakrishna; PARIKH, Devi et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, s. 618–626. ISBN 978-1-5386-1032-9.

- [53] ROSEBROCK, Adrian. *OpenCV Connected Component Labeling and Analysis*. Online. PyImageSearch. Dostupné z: <https://pyimagesearch.com/2021/02/22/opencv-connected-component-labeling-and-analysis/> [cit. 2025-12-11].
- [54] KINGMA, Diederik P.; Ba, Lei Jimmy. Adam: A Method for Stochastic Optimization. Online, preprint. In: *3rd International Conference on Learning Representations, ICLR 2015*. 2015. Dostupné z: <https://arxiv.org/pdf/1412.6980>.
- [55] LIN, Min; CHEN, Qiang a YAN, Shuicheng. *Network in network*. Online, preprint. arXiv. Dostupné z: <https://arxiv.org/pdf/1312.4400>[cit. 2025-01-28].
- [56] LIVIERIS, Ioannis E.; PINTELAS, Emmanuel; KIRIAKIDOU, Niki a PINTELAS, Panagiotis. Explainable image similarity: integrating siamese networks and Grad-CAM. *Journal of Imaging*. 2023, vol. 9, no. 10, article 224. ISSN 2313-433X.
- [57] AUBIN, Verónica; MORA, Marco a SANTOS-PEÑAS, Matilde. Off-line writer verification based on simple graphemes. *Pattern Recognition*. 2018, vol. 79, s. 414–426. ISSN 0031-3203.
- [58] COMPUTER VISION LAB. *CVL-Database*. Online. CVL TU Wien. Dostupné z: <https://cvl.tuwien.ac.at/research/cvl-databases/an-off-line-database-for-writer-retrieval-writer-identification-and-word-spotting/>. [cit. 2025-05-09].
- [59] CSAFE. *Handwriting Database*. Online. Csafe. Dostupné z: <https://data.csafe.iastate.edu/HandwritingDatabase/>. [cit. 2025-05-09].
- [60] KATHURIA, Ishani. *Handwritten Signature Datasets* . Online. Kaggle. Dostupné z <https://www.kaggle.com/datasets/ishanikathuria/handwritten-signature-datasets>. [cit. 2025-05-09].

Seznam symbolů a zkratek

Adam	Adaptive moment estimation
AUC	<i>Area Under Curve</i>
CNN	Convolutional Neural Network
DNN	Deep Neural Network
FN	False Negative
FP	False Positive
FPR	<i>False Positive Rate</i>
GMM	Gaussian Mixture Model
Grad-CAM	Gradient-weighted Class Activation Mapping
HMM	Hidden Markov Model
<i>k</i>-NN	<i>k</i> -Nearest Neighbors
LBP	Local Binary Pattern
MBConv	Mobile inverted Bottleneck Convolution
MHSA	Multi-Head Self-Attention
MLP	Multi-Layer Perceptron
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
SE	Squeeze and Excitation
SIFT	Scale-Invariant Feature Transform
SNN	Shallow Neural Network
SVM	Support Vector Machines
TN	True Negative
TNR	<i>True Negative Rate</i>
TP	True Positive

TPR *True Positive Rate*
ViT Vision Transformer

A Obsah elektronické přílohy

Vzhledem k omezené velikosti elektronické přílohy se zde nachází pouze soubor `gitlab_link.pdf` obsahující odkaz na veřejný GitLab repositář. Tento repositář obsahuje skripty určené k trénování a testování modelů. Dále pak výstupy v podobě evaluačních metrik. S výjimkou datových množin a teplotních map jsou zveřejněny všechny materiály. Struktura repositáře je následující.

```
k_verification/ ..... kořenový adresář repositáře
├── test_models ..... testování modelů
│   ├── CNN_test ..... testování CNN
│   │   ├── CNN_datasets
│   │   ├── CNN_models
│   │   ├── CNN_results
│   │   └── CNN_test.py
│   ├── SCNN_avg_test ..... testování siamské sítě s autoenkodérem
│   │   ├── SCNN_avg_datasets
│   │   ├── SCNN_avg_models
│   │   ├── SCNN_avg_results
│   │   └── SCNN_avg_test.py
│   └── SCNN_single_test ..... testování konvoluční siamské sítě
│       ├── SCNN_single_datasets
│       ├── SCNN_single_models
│       ├── SCNN_single_results
│       └── SCNN_single_test.py
├── train_models ..... trénování modelů
│   ├── CNN_all ..... trénování CNN
│   │   ├── CNN_a_dataset_BP
│   │   ├── CNN_all_dataset_BP
│   │   ├── CNN_k_dataset_BP
│   │   ├── CNN_results
│   │   └── CNN.py
│   ├── SCNN_avg_all ..... trénování siamské sítě s autoenkodérem
│   │   ├── SCNN_a_dataset_BP
│   │   ├── SCNN_k_dataset_BP
│   │   ├── SCNN_avg_results
│   │   └── SCNN_avg.py
│   └── SCNN_single_all ..... trénování konvoluční siamské sítě
│       ├── SCNN_single_chars_BP
│       ├── SCNN_single_results
│       └── SCNN_single.py
├── README.md
└── requirements.txt
```