

Comparison of different types of neural networks for predicting the suitability of the produced agglomerate.

1st Čarnická Tatiana

*Fakulta elektrotechniky a informatiky
Technická Univerzita v Košiciach*
tatiana.carnicka@student.tuke.sk

2nd Firda Dávid

*Fakulta elektrotechniky a informatiky
Technická Univerzita v Košiciach*
david.firda@student.tuke.sk

3rd Haraksim Daniel

*Fakulta elektrotechniky a informatiky
Technická Univerzita v Košiciach*
daniel.haraksim@student.tuke.sk

4th Sirotnák Jonáš

*Fakulta elektrotechniky a informatiky
Technická Univerzita v Košiciach*
jonas.sirotnak@student.tuke.sk

5th Skopár Lukáš

*Fakulta elektrotechniky a informatiky
Technická Univerzita v Košiciach*
lukas.skopar@student.tuke.sk

6th Ing. Stredanský Maroš

*Fakulta elektrotechniky a informatiky
Technická Univerzita v Košiciach*
maros.stredansky@student.tuke.sk

Abstract—Táto štúdia sa zameriava na porovnanie výkonnosti rôznych typov neurónových sietí pri predikcii hodnoty rozsevu pri výrobe aglomerátu v oceliarskom priemysle. Konkrétne sa analyzujú LSTM, GRU a CNN siete vo vzťahu k predikcii hodnôt rozsevu aglomerátu. Výsledky poskytujú dôležité poznatky o schopnostiach týchto modelov zachytávať časové závislosti a efektívne spracovávať dátové sekvencie. Táto práca prispieva k pochopeniu využitia moderných metód strojového učenia na optimalizáciu výrobných procesov v oceliarskom priemysle.

Index Terms—neural networks, LSTM, CNN, GRU, prediction, recurrent neural networks, data processing

ÚVOD

V dnešnom priemyselnom prostredí je kľúčové zameranie sa na optimalizáciu výrobných procesov, minimalizáciu nákladov a zlepšovanie kvality produktov. Jedným z významných procesov v oceliarskom priemysle je výroba aglomerátu, ktorá má výrazný dopad na efektívnosť a ekonomiku celého výrobného cyklu. Jedným z kritických aspektov, ktorý určuje kvalitu aglomerátu, je jeho granulometrické zloženie, teda rozsev. Tradičné metódy predikcie a monitorovania rozsevu sa stretávajú s problémami spojenými s variabilitou surovín a neustálymi zmenami v podmienkach výroby.

Táto práca sa zaoberá využitím moderných metód strojového učenia, konkrétne sietí LSTM (Long Short-Term Memory) [1], GRU [2] (Gated Recurrent Unit), a konvolučných neurónových sietí (CNN) [3], na predikciu hodnoty rozsevu aglomerátu. Siete LSTM a GRU sú výnimočné vďaka svojej schopnosti modelovať časové závislosti a zachytiť dynamiku sekvenčných dát, čo ich robí ideálnymi pre analýzu časovo variabilných výrobných procesov. CNN, na druhej strane, sú využívané pre svoju schopnosť efektívne spracovať štruktúrované dátové postupnosti a extrahovať relevantné vlastnosti z viacrozmerných dátových polí.

Táto štúdia sa zameriava na evaluáciu a porovnávanie účinnosti modelov LSTM, GRU a CNN v predikcii granulo-

metrického rozloženia počas výrobného procesu aglomerátu. Zároveň poskytuje analýzu, ako môže začlenenie týchto sofistikovaných strojových učiacich sa metód do priemyselných operácií vylepšiť procesné rozhodovanie a zvýšiť efektivitu výrobných procesov.

I. DÁTA

Dáta z produkcie aglomerátu v spoločnosti US Steel Košice sú kľúčovým zdrojom informácií o výrobnom procese aglomerátu. Tento dataset obsahuje rôzne údaje získané zo senzorov umiestnených na špekacom páse a tiež hodnoty rozsevu v rozsahu 0-5 milimetrov. Tieto merania boli vykonávané v 30-minútových intervaloch po dobu štyroch rokov na jednom špekacom páse.

Je dôležité zdôrazniť, že v kontexte výroby aglomerátu má hodnota rozsevu významný vplyv na kvalitu výrobku. Ak je rozsev menší ako 12, aglomerát je považovaný za produkt dobrej kvality. Naopak, hodnoty rozsevu nad týmto prahom znižujú kvalitu aglomerátu, čo negatívne ovplyvňuje konečný produkt a môže viesť k problémom v ďalších fázach výrobného procesu.

Tieto údaje sú neoceniteľné pre sledovanie a optimalizáciu výrobného procesu, keďže umožňujú analyzovať a predpovedať vývoj produkcie a identifikovať potenciálne anomálie alebo slabé stránky v procese.

A. Predspracovanie dát

Naša práca na predspracovaní údajov zahŕňa série transformácií na dvoch datasetoch. Prvý dataset (DT1) zahŕňa záznamy o aglomeráte zaznamenané každých 30 minút, zatiaľ čo druhý dataset (DT2) obsahuje merania hodnoty rozsevu vykonané manuálne o 11:00 a 23:00. Cieľom je tieto údaje pripraviť pre analýzu a modelovanie. Začíname konverziou časových intervalov na štandardné dátumové formáty a ich zarovnaním do požadovaných časových intervalov.

V nasledujúcom kroku sme upravili časové intervaly v DT2, nastavili sme ich na špecifické hodiny a ak bolo potrebné, posunuli ich na predchádzajúci deň. Následne sme v DT1 vyčistili neplatné a irelevantné údaje.

Na záver sme oba datasety zlúčili na základe zhodných časových intervalov a odstránili sme z výsledného datasetu všetky nepotrebné stĺpce. Tieto úpravy zabezpečujú, že dáta sú správne pripravené na ďalšie analýzy a modelovanie pomocou neurónovej siete.

B. Analýza dát

Pri preskúvaní distribúcie hodnôt "Rozsev 0-5" je vidieť, že hodnoty sa väčšinou sústreďujú okolo strednej hodnoty a vykazujú normálne rozdelenie s maximom približne pri hodnote 10. Avšak, zaznamenali sme aj prítomnosť extrémnych hodnôt, ktoré môžu dosahovať hodnoty až do 40. Tieto odchýlené hodnoty pramenia z určitých špecifických podmienok, teda sú ovplyvnené konkrétnymi faktormi vo výrobnom procese.

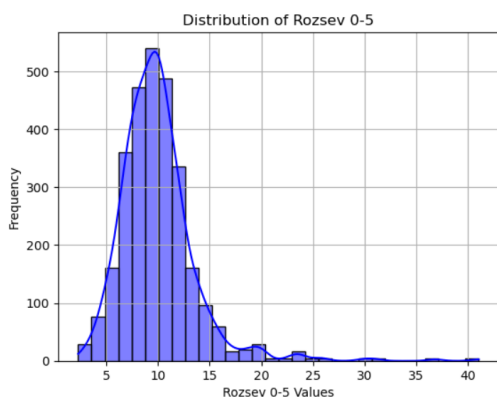


Fig. 1. Distribúcia spracovaných dát.

Tieto poznatky sú podstatné pre určenie relevancie a presnosti získaných dát, čo je kľúčové pre ich efektívne využitie v ďalších krokoch.

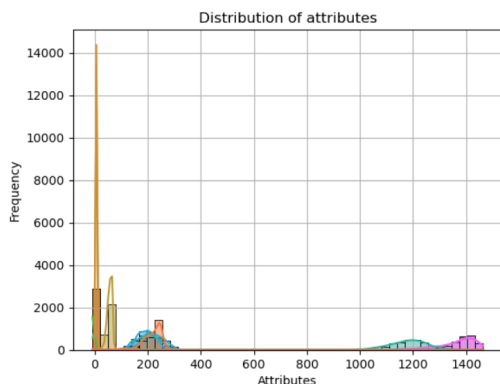


Fig. 2. Distribúcia spracovaných dát.

Ďalší graf ukazuje výkyvy, ktoré môžu zodpovedať určitým vonkajším vplyvom alebo vnútorným procesom, ktoré ovplyvňujú cieľovú premennú. Môže ísť o dáta zaznamenávané

v rôznych časových obdobiach, kde vysoké špičky môžu indikovať mimoriadne udalosti alebo zmeny v podmienkach.

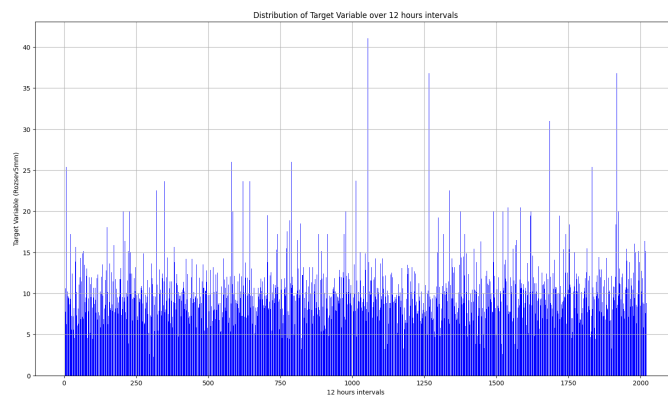


Fig. 3. Distribúcia cieľovej hodnoty v 12 hodinových intervaloch.

Je dôležité poznamenať, že dáta sú rozdelené do približne rovnakých intervalov, čo naznačuje, že merania boli vykonávané systematicky a konzistentne. Tento graf môže posloužiť na identifikáciu trendov, cyklov alebo na ďalšie štatistické analýzy zamerané na pochopenie dynamiky cieľovej premennej v zadanom časovom rozmedzí.

II. PREDIKCIA HODNÔT POMOCOU UMELEJ INTELIGENCIE

Predikcia hodnôt pomocou umelej inteligencie predstavuje revolučný prístup k analýze a predpovedaniu v rôznych odvetviach priemyslu. Využitie moderných techník strojového učenia, ako sú LSTM, GRU a CNN siete, umožňuje presnejšie a efektívnejšie modelovanie časových radov a komplexných dátových vzorov. Tento pokročilý prístup k predikcii hodnôt výrazne prispieva k optimalizácii výrobných procesov a zvyšovaniu efektivity priemyselných operácií. V nasledujúcej časti sa bližšie pozrieme na výhody a výzvy spojené s využitím umelej inteligencie na predikciu hodnôt v priemyselných aplikáciách, a na samotné typy LSTM, GRU a CNN sieti.

A. LSTM model

LSTM alebo Dlhodobá krátkodobá pamäť, je pokročilý typ rekurentnej neurónovej siete, ktorý vyvinuli Hochreiter a Schmidhuber [1]. Táto architektúra je špeciálne navrhnutá na modelovanie sekvenčných dát a je výnimočná v zachytávaní dlhodobých závislostí v dátových sekvenciách. LSTM sa často využíva v aplikáciách, ktoré spracovávajú časové rady a iné typy sekvencií.

Tradičné RNN [4] čelia problémom pri učení dlhodobých závislostí kvôli ich štruktúre s jediným skrytým stavom, ktorý sa aktualizuje v čas, čo komplikuje učenie. LSTM rieši tento problém prostredníctvom špeciálnej pamäťovej bunky, ktorá umožňuje uchovávanie informácií na dlhšie časové obdobie. Týmto spôsobom sa môžu siete LSTM efektívne učiť dlhodobé závislosti v sekvenčných dátach.

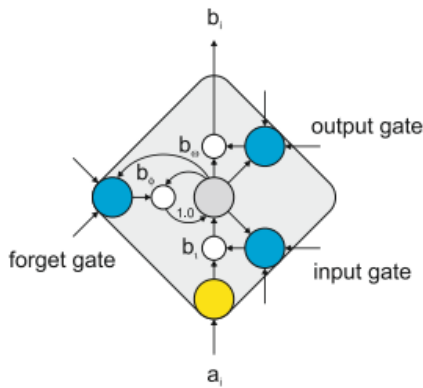


Fig. 4. LSTM pamäťová bunka s bránami [5].

Pamäťová bunka v LSTM je kontrolovaná tromi bránami [6], [7]: vstupnou bránou (input gate), zabudlivou bránou (forget gate) a výstupnou bránou (output gate). Tieto brány umožňujú selektívne rozhodnúť, ktoré informácie budú pridané do pamäťovej bunky, ktoré z nej budú odstránené a ktoré budú poslané ďalej. Tento mechanizmus poskytuje LSTM schopnosť efektívne spracovať sekvenčné dáta, udržiavať alebo eliminovať informácie podľa potreby. Veľkosť okna v modeloch LSTM určuje, koľko predchádzajúcich krokov (časových intervalov) model zohľadňuje pri učení závislostí v dátach.

1) *Prečo práve LSTM:* Pri výbere modelu pre predikciu hodnoty Rozsevu je kľúčové zvážiť niekoľko faktorov, ako je štruktúra dát, charakteristika časového radu a požiadavky na presnosť predikcií.

LSTM modely sú často vhodnou voľbou pre takéto úlohy z niekoľkých dôvodov:

- **Zachytenie dlhodobých závislostí:** LSTM siete majú schopnosť udržiavať a aktualizovať informácie z minulosti v dlhobehkej pamäti, čo je dôležité pri predikcii hodnôt v časových radoch.
- **Spracovanie sekvenčných dát:** LSTM modely efektívne spracovávajú sekvenčné dáta, ako sú časové rady, a sú navrhnuté na učenie sa z sekvencií dát a identifikáciu vzorov.
- **Flexibilita:** LSTM modely majú nastaviteľné parametre, čo umožňuje ich prispôbenie konkrétnym požiadavkám a charakteristikám dát.

Tieto vlastnosti robia z LSTM modelov dobrého kandidáta na predikciu hodnôt Rozsevu. Je však dôležité nájsť správny model, aby sa dosiahla optimálna predikčná presnosť.

2) *Výhody a nevýhody:*

- **Výhody [8]:** Siete LSTM dokážu efektívne zachytiť dlhodobé závislosti v dátach vďaka pamäťovej bunke, ktorá uchováva informácie dlhodobo a rieši problémy s miznúcimi gradientmi.
- **Nevýhody [8]:** Napriek ich schopnostiam majú siete LSTM vyššiu výpočtovú náročnosť v porovnaní s jednoduchšími modelmi. Ich tréning je tiež časovo

náročnejšie a vyžaduje väčšie množstvo dát a dlhší čas tréningu na dosiahnutie vysokého výkonu.

3) *Rozdelenie dát pre vstup do modelu:* Na začiatku sa vstupný dataset konvertuje na typ float pre numerickú konzistenciu. Odstránením hodnôt NaN zabezpečíme integritu dát. Následne rozdelíme dáta na vstupné a cieľové hodnoty, a vytvoríme tréningovú množinu (80% dát) a dočasnú množinu (zvyšných 20%). Tú ďalej rozdelíme na validačnú a testovaciu množinu v pomere 1:1.

Všetky množiny sa normalizujú pomocou MinMaxScaler, čím sa hodnoty upravujú do rozsahu -1 až 1, čo je kľúčové pre neurónové siete a iné algoritmy strojového učenia.

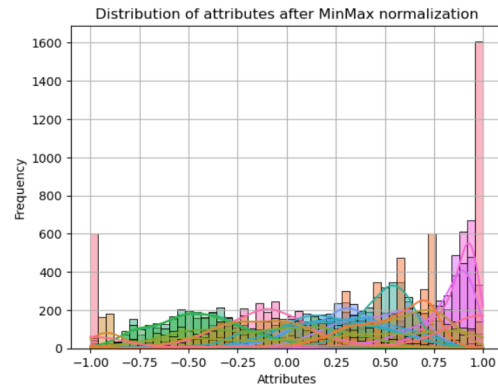


Fig. 5. Distribúcia dát po normalizácii.

Nakoniec sa tieto množiny transformujú na tenzory pre kompatibilitu s neurónovými sieťami, pričom upravíme dátový typ na np.float32 a tvar cieľových tenzorov na (batch_size, 1). Tréningové a validačné množiny rozdelíme pomocou DataLoader do dávok pre efektívnejšie spracovanie dát počas tréningu a validácie, čo zlepšuje výpočtovú efektívnosť a stabilitu modelu.

4) *Štruktúra a parametre modelu:* Naš LSTM model, vytvorený v PyTorch [9], je zameraný na spracovanie sekvenčných dát a obsahuje LSTM vrstvu a plne prepojené vrstvy (FC) pre predikcie.

Funkcia forward implementuje tok dát modelom inicializáciou skrytých stavov LSTM, ktoré prechádzajú cez LSTM vrstvu s parametrami: počet vstupov, veľkosť pamäte a dropout. Toto pokračuje cez FC vrstvy s aktiváciou relu a lineárnou aktiváciou na výstupe, kde prvá vrstva má 64 neurónov, druhá vrstva má 32 neurónov, tretia vrstva má 16 neurónov a posledná má 1 neurón pre predikciu na výstupe.

Parametre modelu zahŕňajú vstupnú dimenziu podľa počtu stĺpcov v dátach, skrytú dimenziu 150, a 3 LSTM vrstvy pre hĺbku siete. Inicializovali sme model s dropout 0,1 proti pretrénovaniu, používame MSE na meranie chýb a Adam optimalizátor s learning rate 0,005 pre efektívne tréningovanie.

B. CNN model

Konvolučné neurónové siete (CNN) [10], [11] predstavujú široko používaný typ neurónových sietí v oblasti počítačového videnia, kde excelujú vo spracovaní obrázkových dát. Avšak

ich aplikácia sa neobmedzuje len na túto oblasť. My sme otestovali použitie CNN na predikciu veľkosti rozsevu aglomerátu. Pre dôkladné preskúmanie tejto možnosti sme vytvorili dva rozdielne modely.

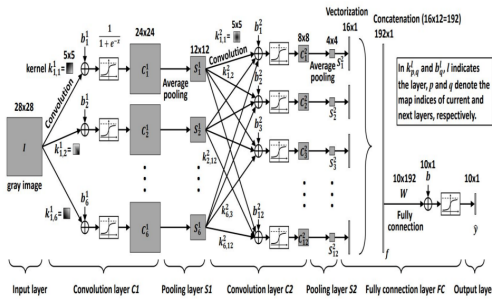


Fig. 6. Príklad jednoduchšej konvulcnej neurónovej siete [10].

1) *Prečo práve CNN:* Pre našu úlohu predikcie veľkosti rozsevu aglomerátu sme sa rozhodli otestovať aj CNN kvôli jej schopnosti efektívne pracovať s časovými radmi a zachytávať vzory v sekvenciálnych dátach. Konvulčné vrstvy sú schopné detekovať priestorové vzory vstupných dát, čo je v prípade našej úlohy dôležité, pretože veľkosť rozsevu aglomerátu môže byť ovplyvnená rôznymi faktormi v procese výroby. Navyše, distribúcia hodnôt rozsevu aglomerátu vykazuje extrémne hodnoty, ktoré by mohli byť efektívne zachytené a predikované pomocou konvulčných vrstiev.

2) *Výhody a nevýhody:*

- **Výhody:**

Vďaka svojej architektúre dokážu extrahovať kľúčové detaily z obrázkov, videí a iných dát, postupne od jednoduchých tvarov k zložitým vzťahom [12]. Taktiež sú efektívne paralelizovateľné, čo umožňuje ich rýchle tréningovanie a predikciu aj na moderných grafických kartách (GPU). CNN zdieľajú váhy medzi neurónmi vo filtroch, čo umožňuje efektívne učenie aj na malých datasetoch a zabraňuje preučeniu.

- **Nevýhody:**

Vyžadujú veľké množstvá tréningových dát, čo môže byť pri niektorých aplikáciách obmedzujúce. Hlboké CNN sú náchylné na pretrénovanie a interpretácia výsledkov môže byť kvôli ich zložitosti náročnejšia [13]. Ak sú v tréningových dátach niektoré vzory zriedkavé alebo neexistujú, môže byť obtiažne naučiť CNN ich správne rozpoznávať.

3) *Rozdelenie dát pre vstup do modelu:* V oboch modeloch začíname načítaním veľkého datasetu a následným zmiešaním a transformáciou dát, kde indexy riadkov sú náhodne zamiešané a vytvára sa nový dataframe, kde sú dáta zoradené podľa týchto zmiešaných indexov. Potom nasleduje čistenie dát, kde sa odstránia riadky s chýbajúcimi hodnotami. Ďalej sa dáta normalizujú pomocou MinMaxScaler, čo znamená, že sa škálujú tak, aby každý atribút mal hodnoty v rozsahu medzi 0 a 1. Nakoniec sa dáta rozdelia do tréningovej, testovacej a validačnej sady. Testovacia sada tvorí 20% z celkového množstva dát, zatiaľ čo tréningová a validačná sada sa ďalej rozdeľujú s pomerom 70:30, kde 70% tvorí tréningová a 30% validačná sada.

4) *Štruktúra a parametre modelu:* Sieť **Modelu č. 1** sa skladá z troch konvulčných vrstiev, z ktorých každá extrahuje lokálne vzory zo vstupných dát. Počet filtrov sa postupne znižuje z 128 na 64 a 32, čím sa sieť zameriava na najvýznamnejšie črty v neskorších fázach spracovania. Normalizačné vrstvy BatchNormalization stabilizujú a urýchľujú proces učenia, zatiaľ čo aktivačná funkcia LeakyReLU [14] umožňuje sieti zachytiť aj zložité nelineárne vzťahy v dátach. Vrstvy MaxPooling redukujú rozmery výstupu konvulčných vrstiev, čím znižujú výpočtové nároky a zabraňujú pretrénovaniu siete. Nakoniec, plne prepojené vrstvy integrujú informácie z predchádzajúcich vrstiev a vykonávajú konečné klasifikačné rozhodnutie o veľkosti rozsevu aglomerátu. **Modeli č. 2** začína vytvorením sekvenčného modelu, ktorý je základným kontajnerom pre rôzne typy vrstiev v Keras. Následuje prídanie vstupnej vrstvy s určeným tvarom na základe počtu funkcií (atribútov) vstupných dát. Model využíva štyri konvulčné vrstvy na detekciu vzorov vo vstupných dátach. Tieto vrstvy majú nastavený rôzny počet filtrov (32, 64, 128, 64). Každá konvulčná vrstva využíva aktivačnú funkciu ReLU. Po konvulčných vrstvách nasledujú vrstvy max-pooling, ktoré slúžia na znižovanie dimenzie výstupu z konvulčných vrstiev. Ďalej je pridaná vrstva Dropout, ktorá slúži na reguláciu overfittingu tým, že náhodne vypne určité neuróny počas tréningovania s určitou pravdepodobnosťou. Nakoniec je pridaná plne prepojená vrstva s 10 neurónmi a aktiváciou ReLU, ktorá slúži na vytvorenie výstupu modelu.

C. GRU model

Ako ďalší model sme sa rozhodli implementovať Gated Recurrent Unit (GRU). Skladá sa z dvoch brán, resetovacej a aktualizacej brány. Resetovacia brána určuje ako sa bude kombinovať aktuálny vstup v stave s historickou pamäťou. Aktualizačná brána je zodpovedná ako veľmi by mala byť pamäť udržiavaná v uzle [15].

1) *Prečo práve GRU:* Pre GRU sme sa rozhodli preto, lebo je to druh rekurentnej neuronovej siete, podobný modelu LSTM. Narozdiel od LSTM je GRU rýchlejší a vyžaduje menej dát a parametrov [16]. Daný model sme si zvolil, lebo je vhodný na data v určitom časovom priebehu, čo vyhovuje našim nasnímaným dátam v časových intervaloch a slúži ako vhodná alternatíva pre porovnanie výsledkov s LSTM modelom.

2) *Výhody a nevýhody:*

- **Výhody:**

Veľkou výhodou GRU modelov je ich efektívne a procesovo nenáročné tréningovanie vďaka ich jednoduchšej architektúre. Vyžadujú menej parametrov a dát oproti modelom ako napríklad CNN a LSTM. Taktiež sú efektívne pre sekvenčné data, ako tie naše, a na úlohy predikcie.

- **Nevýhody:**

Na druhej strane je ich jednoduchá architektúra limitujúca a nedokáže zachytiť viac komplexnejšie vzory, majú limitovanú dlhodobú pamäť a sú menej flexibilné kvôli ich mechanizmu brán.

3) *Rozdelenie dát pre vstup do modelu:* Vstupné data očistili od nečíselných a duplikátnych hodnôt. Pri implementovaní datasetu sme si ho rozdelili na vstupné data, čo reprezentovalo 17 stĺpcov dát a výstupné data, data, ktoré budeme predikovať a to hodnotu rozsevu pod 5 milimetrov. Následne sme ich škálovali aby sme zaistili stabilitu modelu. Vstupné a výstupné dátá sme rozsekli na dáta tréningové, validačné a testovacie a to v pomere 80:14:6.

Keďže náš model je implementovaný pomocou rozhrania Tensorflow [17], tak sme potrebovali naše dáta pretransformovať na tenzory.

4) *Štruktúra a parametre modelu:* Model je sekvenciálna neurónová sieť pozostávajúca z dvoch vrstiev GRU, nasledovaných viacerými plne prepojenými vrstvami. Prvá vrstva GRU má 128 jednotiek a je nastavená na návrat sekvencií, zatiaľ čo druhá vrstva GRU má 64 jednotiek a vracia iba výstup posledného časového kroku. Obe vrstvy GRU používajú aktivačnú funkciu Rectified Linear Unit (ReLU) na zavedenie nelinearity. Po vrstvách GRU nasledujú štyri plne prepojené vrstvy s 64, 32, 16 a 8 jednotkami, pričom každá používa tiež aktivačnú funkciu ReLU. Tretia a štvrtá plne prepojená vrstva majú aplikovanú regularizáciu L2 na ich maticu jadier na minimalizáciu preučenia. Po tretej hustej vrstve je pridaná vrstva Dropout s mierou 0,1 na ďalšiu prevenciu preučenia tým, že náhodne vypúšťa 10 percent vstupných dát počas tréningovania. Posledná výstupná vrstva pozostáva z jediného hustého neurónu, ktorý predstavuje predikciu modelu pre regresné úlohy, bez špecifikovanej aktivačnej funkcie, čo naznačuje lineárnu aktiváciu (identitnú funkciu) pre predikciu spojitého výstupu. Táto architektúra je navrhnutá pre sekvencie dát s 17 príznakmi na každom časovom kroku, s cieľom predikovať jediný spojitý výstup.

III. POROVNANIE MODELOV

Pri porovnávaní modelov LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit) a CNN (Convolutional Neural Network) v kontexte predikcie časových radov, je dôležité zvážiť ich unikátne charakteristiky a výkonnosť.

A. Výsledky LSTM modelu

TABLE I
SUMARIZÁCIA LSTM MODELU

Parameter	Hodnota
num_epochs	900
hidden_dim	150
layer_dim	3
learning_rate	0.005
dropout	0.1
optimizer	Adam
Vyhodnotenie modelu:	
Training Loss	0.8142
Validation Loss	0.2597
Výsledky na valid. množine:	
MSE	0.2640
MAE	0.2332
R-squared	0.9800
Custom Accuracy	0.9282

Z grafu je zrejmé, že obidve krivky strát rýchlo klesajú v úvodných fázach tréningovania, čo naznačuje, že model efektívne sa učí z dát. Po tomto rýchlom poklese, krivky strát sa stabilizujú a vykazujú len malé výkyvy.

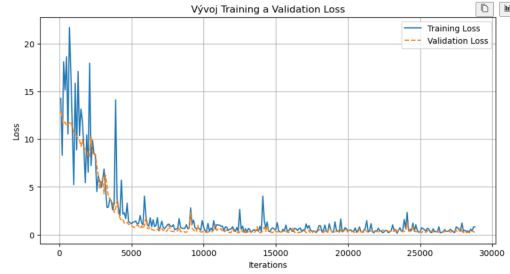


Fig. 7. Vývoj tréningových a validačných chýb.

B. Výsledky CNN modelu

TABLE II
SUMARIZÁCIA CNN MODELU

Parameter	Hodnota
num_epochs	300
kernel_size	3
layer_dim	3
learning_rate	0.01
dropout	0.5
optimizer	SGD
Vyhodnotenie modelu:	
Training Loss	1.0870
Validation Loss	2.0825
Výsledky na valid. množine:	
MSE	2.6402
MAE	0.9159
R-squared	0.8361
Custom Accuracy	0.8331

Pre vyhodnotenie sme sa rozhodli použiť zmieneny model č. 1.

C. Výsledky GRU modelu

TABLE III
SUMARIZÁCIA GRU MODELU

Parameter	Hodnota
num_epochs	500
layer_dim	3
learning_rate	0.001
dropout	0.1
optimizer	Adam
Vyhodnotenie modelu:	
Training Loss	0.2281
Validation Loss	4.8748
Výsledky na valid. množine:	
MSE	2.7556
MAE	1.1123
R-squared	0.81340
Custom Accuracy	0.7707

IV. VYHODNOTENIE

LSTM [1] modely sú známe svojou schopnosťou efektívne spracovávať dlhodobé závislosti v dátach vďaka svojim pamäťovým bunkám a bránam, ktoré regulujú tok informácií. Na druhej strane, GRU [2], ktoré je variantom LSTM, má jednoduchšiu štruktúru, keďže kombinuje vstupné a zabudlivé brány do jednej brány aktualizácie. Táto zjednodušená štruktúra umožňuje rýchlejšie trénovanie a efektívnejšie spracovanie, pričom stále zachováva vysokú úroveň výkonnosti. CNN modely [3], tradične používané v analýze obrazov a videí, sa ukázali byť účinné aj v predikcii časových intervalov, najmä vďaka svojej schopnosti detekovať vzťahy a závislosti v dátach. Ich konvolyčné vrstvy efektívne spracovávajú vstupné dáta na extrahovanie dôležitých vlastností, ktoré sú následne využité v plne prepojených vrstvách pre finálne predikcie.

Najlepší model v tomto porovnaní bol LSTM model. Po trénoch a validácii dosiahol nízku hodnotu trénovacej straty (0.8142) a ešte nižšiu hodnotu validačnej straty (0.2597). Tento model dosiahol priemernú kvadratickú chybu (MSE) na validačnej množine 0.2640 a priemernú absolútnu chybu (MAE) 0.2332. R-kvadrátový koeficient pre tento model bol vysoký, a to 0.9800, a presnosť modelu bola 0.9282. Tieto výsledky naznačujú, že LSTM model dosiahol najlepšiu schopnosť predikcie hodnoty rozsevu aglomerátu a je najlepším modelom v tomto kontexte.

ZÁVER

Na záver našej štúdie sme porovnali modely LSTM, GRU a CNN v predikcii hodnoty rozsevu v procese výroby aglomerátu. Model LSTM, s jeho schopnosťou efektívne spracovávať dlhodobé závislosti v dátach, sa ukázal byť najúspešnejší. Výsledky ukazujú, že LSTM dosiahol nielen nízku trénovaciu stratu 0.8142 a validačnú stratu 0.2597, ale aj vysokú presnosť predikcií s hodnotou MSE 0.2640, MAE 0.2332, R-squared 0.9800 a presnosťou 0.9282.

Naopak, GRU modely, aj keď sú efektívnejšie v niektorých scenároch vďaka jednoduchšej štruktúre a rýchlejšiemu trénovaniu, nedosiahli takú presnosť ako LSTM. CNN modely, tradične silné v analýze obrazových a štruktúrovaných dát, preukázali schopnosť efektívne spracovávať a extrahovať vzťahy v časových intervaloch, ale ich výsledky boli mierne slabšie ako pri modeli LSTM.

Tieto poznatky poukazujú na to, že pre komplexné a dynamické predikcie, kde sú dôležité dlhodobé závislosti, je LSTM najvhodnejším modelom. Využitie LSTM v praxi by mohlo výrazne prispieť k lepším rozhodnutiam a zefektívneniu výrobných procesov, najmä v oceliarskom priemysle, kde je kvalita a predikcia rozsevu aglomerátu kľúčová. Naša štúdia teda potvrdzuje významný potenciál LSTM sietí pre zlepšenie predikcie a optimalizáciu procesov v rôznych priemyselných aplikáciách.

REFERENCES

- [1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [2] M. Habiba and B. Pearlmutter, "Neural ordinary differential equation based recurrent neural network model," 05 2020.

- [3] M. Al-Zuhairi, B. Pradhan, H. Shafri, and H. Hamid, *Applications of Deep Learning in Severity Prediction of Traffic Accidents*, 01 2019, pp. 793–808.
- [4] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *ArXiv*, vol. abs/1808.03314, 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:51968707>
- [5] M. Sundermeyer, R. Schlüter, and H. Ney, "Lstm neural networks for language modeling," in *Interspeech*, 2012. [Online]. Available: <https://api.semanticscholar.org/CorpusID:18939716>
- [6] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, Oct 2000.
- [7] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with lstm recurrent networks," *Journal of Machine Learning Research*, vol. 3, pp. 115–143, 2002.
- [8] Aakarshachug, "Deep learning: Introduction to long short-term memory," <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>, 2023, accessed: 2024-05-06.
- [9] PyTorch, "torch.nn.LSTM documentation," <https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html>, 2023, accessed: 2024-05-06.
- [10] P. Sincák, *Neurónové siete - inžiniersky prístup*, 4th ed. TUKE, June 2023.
- [11] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*. IEEE, 2017, pp. 1–6.
- [12] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.
- [13] L. Alzubaidi et al., "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, pp. 1–74, 2021.
- [14] C. Nwankpa et al., "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378*, 2018.
- [15] A. Sethia and P. Raut, "Application of lstm, gru and ica for stock price prediction," in *Information and Communication Technology for Intelligent Systems*, S. C. Satapathy and A. Joshi, Eds. Singapore: Springer Singapore, 2019, pp. 479–487.
- [16] M. Pan, H. Zhou, J. Cao, Y. Liu, J. Hao, S. Li, and C.-H. Chen, "Water level prediction model based on gru and cnn," *IEEE Access*, vol. 8, pp. 60 090–60 100, 2020.
- [17] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>