Primljen: 20.11.2024.
Prihvaćen: 18.12.2024.

Stručni rad
UDK: 004.85

Primjena homomorfne kriptografije na modele strojnoga učenja

Application of homomorphic cryptography to machine learning models

[1a,b*]Mária Krajčí, [2a]Vesna Krajčí
[a]Veleučilište u Rijeci, Vukovarska 58, 51000 Rijeka
[b]Tehnički fakultet u Rijeci, Vukovarska 58, 51000 Rijeka
[*]e-mail: [1*]mkrajci@veleri.hr, [2]vkrajci@veleri.hr

Sažetak: *S obzirom na rastuću potrebu za obradom medicinskih podataka uz osiguranje njihove privatnosti, ovaj rad istražuje primjenu homomorfne kriptografije u treniranju i testiranju modela strojnoga učenja na medicinskim slikama, koristeći pristup temeljen na projektu CryptoNets. Homomorfna kriptografija omogućuje obradu kriptiranih podataka bez dekriptiranja, pružajući tako mogućnost treniranja i analize podataka uz očuvanje privatnosti pacijenata. Testiranje modela na tri odvojena skupa medicinskih podataka (ChestX-Det, Public Lung Dataset i JSRT Dataset) pokazalo je da primjena homomorfne kriptografije rezultira samo blagim smanjenjem točnosti modela. Na skupu ChestX-Det, koji predstavlja multiklasifikacijski problem, mjera F1 ostala je na zadovoljavajućoj razini, dok je na skupovima Public Lung Dataset i JSRT Dataset zabilježena nešto niža, ali prihvatljiva razina točnosti. Primijećeno je da su rezultati na JSRT skupu bili slabiji zbog manjega broja primjera što otežava učenje složenijih obrazaca. U radu se zaključuje da homomorfna kriptografija pruža ravnotežu između zaštite privatnosti i točnosti modela otvarajući mogućnosti za sigurnu obradu osjetljivih medicinskih podataka u kontekstu strojnoga učenja.*

Ključne riječi: *homomorfna kriptografija, CryptoNets, strojno učenje*

Abstract: *Because of the growing need for processing medical data while ensuring its privacy, this paper explores the application of homomorphic encryption in testing machine learning models on medical images, using an approach based on the CryptoNets project. Homomorphic encryption enables the processing of encrypted data without decryption, thus allowing training and data analysis while preserving patient privacy. Testing the model on three separate medical data sets (ChestX-Det, Public Lung Dataset and JSRT Dataset) showed that the application of homomorphic encryption results in*

* corresponding co-author

only a slight reduction in model accuracy. On the ChestX-Det dataset, which represents a multiclass classification problem, the F1 score remained at a satisfactory level, while the Public Lung Dataset and JSRT Dataset recorded slightly lower but acceptable accuracy levels. It was noted that the results on the JSRT dataset were poorer due to the smaller number of samples, which makes it difficult to learn more complex patterns. The paper concludes that homomorphic encryption provides a reasonable balance between privacy protection and model accuracy, opening opportunities for secure processing of sensitive medical data in the context of machine learning.

Key words: *homomorphic encryption, CryptoNets, machine learning*

1. Uvod

S porastom količine medicinskih podataka raste potreba za učinkovitom obradom podataka koja omogućava precizniju dijagnostiku i bolji ishod liječenja uz očuvanje privatnosti podataka. Homomorfna kriptografija nudi rješenje jer omogućuje obradu kriptiranih podataka bez dekriptiranja čime štiti privatnost pacijenata. U ovom radu je primjenjena homomorfna kriptografija s modelom strojnoga učenja na medicinskim slikama, a rezultati pokazuju da je moguće postići visoku sigurnost uz zadržavanje točnosti modela. Ovakav pristup ima značajan potencijal u medicinskim područjima u kojima je zaštita privatnosti ključna (Krajčí, 2023.).

2. Homomorfna kriptografija

Homomorfna kriptografija je jedna od naprednih metoda kriptografije koja omogućava izvođenje aritmetičkih i logičkih operacija na kriptiranim podatcima bez potrebe za njihovim prethodnim dekriptiranjem. Na taj način homomorfna kriptografija omogućava sigurnu obradu podataka modelima strojnoga učenja u oblaku ili na udaljenim serverima gdje podatci ostaju zaštićeni tijekom cijelog procesa obrade. Ovo svojstvo omogućuje primjenu homomorfne kriptografije u područjima poput zdravstva, financija i državne sigurnosti pri čemu je nužno zaštititi privatnost korisnika i povjerljivost podataka (Krajčí, 2023.).

2.1. Vrste homomorfne kriptografije

Homomorfna kriptografija obuhvaća djelomičnu, široku i potpunu homomorfnu kriptografiju ovisno o tome koje su aritmetičke operacije podržane nad kriptiranim podacima (Žigrović, 2017.).

Djelomično homomorfna kriptografija (engl. *Partially Homomorphic Encryption*, PHE) podržava samo jedan tip aritmetičke operacije (npr. samo zbrajanje ili samo množenje). Primjena uključuje RSA algoritam koji podržava samo množenje i Goldwasser-Micali shemu koja omogućava samo zbrajanje (Žigrović, 2017.).

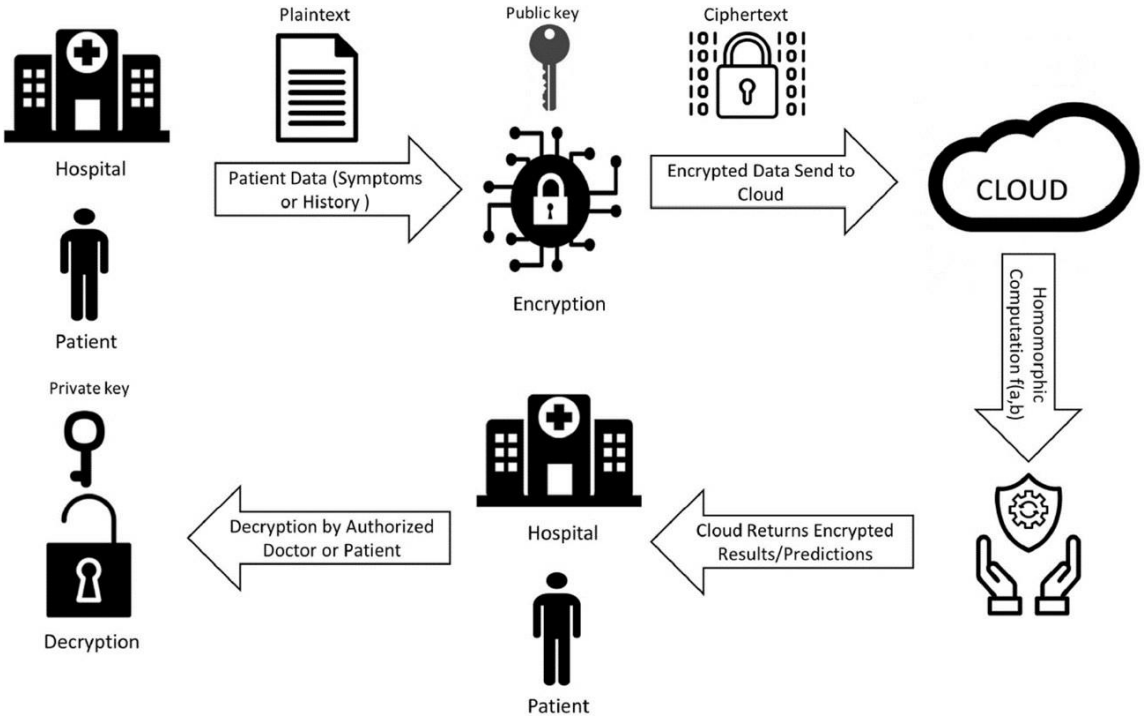
Široka homomorfna kriptografija podržava ograničeni broj aritmetičkih operacija, odnosno kombinacije zbrajanja i množenja, ali s određenim ograničenjem na broj operacija. BGV shema (Brakerski, Gentry i Vaikuntanathan shema) primjer je široko homomorfne kriptografije (Žigrović, 2017.).

Potpuno homomorfna enkripcija (engl. *Fully Homomorphic Encryption,* FHE) omogućuje neograničeni broj aritmetičkih operacija nad kriptiranim podatcima uključujući zbrajanje i množenje što omogućuje složene obrade bez dekriptiranja podataka. Prvi praktični algoritam za potpuno homomorfnu enkripciju razvio je Craig Gentry 2009. godine i od tada je ova metoda brzo napredovala (Žigrović, 2017.).

3. Primjena homomorfne kriptografije

Suvremeni sustavi strojnoga učenja često koriste osjetljive podatke poput zdravstvenih zapisa i sigurnosnih snimaka što stvara izazove za privatnost i usklađenost s propisima. Korištenje potpune homomorfne kriptografije (PHE) omogućuje obradu podataka bez dekriptiranja čuvajući privatnost i sigurnost tijekom analize. Mnoge tvrtke pokreću modele strojnoga učenja putem aplikacijskih programskih sučelja (engl. *Application Programming Interface*, API) na poslužiteljima u oblaku kao na primjeru na slici 1 (obrada medicinskih podataka pomoću modela strojnoga učenja u oblaku) što olakšava pristup računalnoj snazi, ali predstavlja rizik jer pružatelji usluga u oblaku mogu imati pristup nezaštićenim podatcima. PHE uklanja taj rizik omogućujući složene operacije nad kriptiranim podatcima bez kompromitiranja sigurnosti. Iako PHE zahtijeva visoku računalnu snagu i vrijeme obrade, napredak u algoritmima i hardveru povećava učinkovitost čineći ovu metodu sve važnijom za sigurne sustave strojnoga učenja i primjenu u različitim industrijama (Kundan i Bhatia, 2023.).

Slika 1. *Primjena homomorfne kriptografije u medicini.*



Izvor: https://link.springer.com/article/10.1007/s40747-022-00756-z

3.1. Projekt CryptoNets

Projekt CryptoNets obuhvaća neuralne mreže posebno dizajnirane za rad s kriptiranim podatcima omogućujući modelu strojnog učenja izvođenje predikcije nad kriptiranim podatcima bez potrebe za dekriptiranjem. Na taj način povjerljivost podataka ostaje zaštićena jer treća strana (model) nema pristup ključevima potrebnim za dekriptiranje (Downlin et al., 2016.).

Primjer primjene CryptoNets modela je prepoznavanje brojeva u skupu podataka MNIST, pri čemu model postiže visoku točnost od 99 % i omogućava do 59000 predviđanja po satu na jednom računalu. Ova se mreža temelji na djelomičnoj homomorfnoj kriptografiji koja podržava određeni broj aritmetičkih operacija nad kriptiranim podacima (Downlin et al., 2016.).

Struktura CryptoNets koristi prilagođene polinomske funkcije za operacije kao što su ponderirani zbroj, maksimalno i prosječno grupiranje te aktivacijske funkcije (sigmoidna i ReLU funkcija). Budući da djelomična homomorfna kriptografija podržava samo zbrajanje i množenje, navedene funkcije se aproksimiraju polinomima kako bi bile primjenjive na kriptiranim podatcima. Maksimalno i prosječno grupiranje u arhitekturama konvolucijskih neuronskih mreža (engl. *Convolutional Neural Network*, CNN) koje su ključne za izdvajanje značajki su aproksimirane polinomnim funkcijama, dok se aktivacijske funkcije prilagođavaju radi zadržavanja što veće točnosti (Downlin et al., 2016.).

Ova arhitektura omogućuje neuronskim mrežama u projektu CryptoNets postići

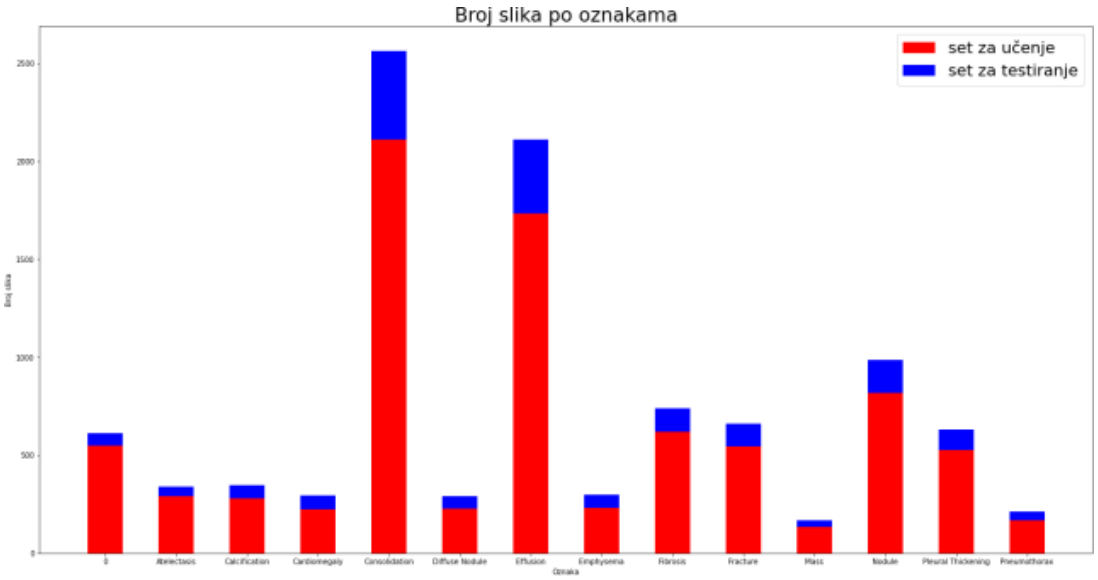
zadovoljavajuću točnost predviđanja uz očuvanje privatnosti čineći ih korisnim rješenjem za primjenu strojnoga učenja na osjetljivim informacijama (Downlin et al., 2016.).

4. Programska implementacija i rezultati rada zasnovanom na CryptoNets

U ovom radu korištena je neuronska mreža trenirana na rendgenskim snimkama, implementirana prema modelu projekta CryptoNets. Postupak uključuje obradu, treniranje i testiranje modela na nekriptiranim slikama te primjenu homomorfne enkripcije na slikama iz skupa za testiranje. Za implementaciju neuronskih mreža i treniranje i testiranje modela korištena je biblioteka Tensorflow, a za kriptiranje i dekripciju biblioteka MS SEAL. Tijek izvođenja algoritma obuhvaća učitavanje slika, treniranje modela, pripremu težina modela za kriptirane slike, kriptiranje skupa za testiranje, predikciju klasa, dekripciju rezultata i usporedbu sa stvarnim vrijednostima.

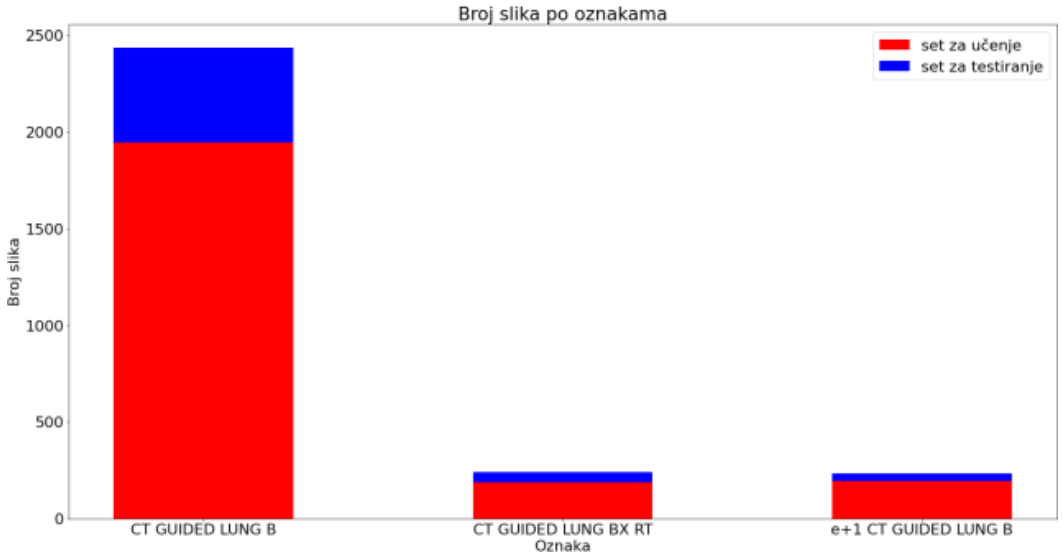
Model je treniran i testiran na tri skupa podataka: ChestX-Det, Public Lung Dataset i JSRT Dataset. ChestX-Det je opsežan skup podataka koji se sastoji od rendgenskih snimaka pluća s različitim dijagnozama plućnih i drugih bolesti poput upale pluća, pleuralnoga izljeva i zatajenja srca čija je podjela slika po klasama prikazana na slici 2. Zbog mogućnosti višeklasne klasifikacije (jedna slika može sadržavati oznake više bolesti) predstavlja izazovni skup za treniranje i evaluaciju modela strojnoga učenja (Lian et al., 2021.). Public Lung Dataset sadrži zbirku javno dostupnih rendgenskih slika pluća koje prikazuju i zdrava pluća i pluća s patološkim promjenama čija je podjela slika po klasama prikazana na slici 3. Skup podataka koristi se za klasifikaciju plućnih bolesti te pruža uvid u osnovne karakteristike zdravih i bolesnih pluća (Reeves et al., 2009.). JSRT Dataset se sastoji od rendgenskih snimaka prsnoga koša s posebnim naglaskom na detekciju čvorova u plućima. Skup sadrži zdrave i bolesne uzorke pluća što ga čini korisnim za detekciju specifičnih abnormalnosti te predstavlja binarnu klasifikaciju (Shiraishi et al., 2000.), a podjela slika po klasama prikazana je na slici 4.

Slika 2. *Raspodjela broja slika po oznakama za skup podataka ChestX-Det.*



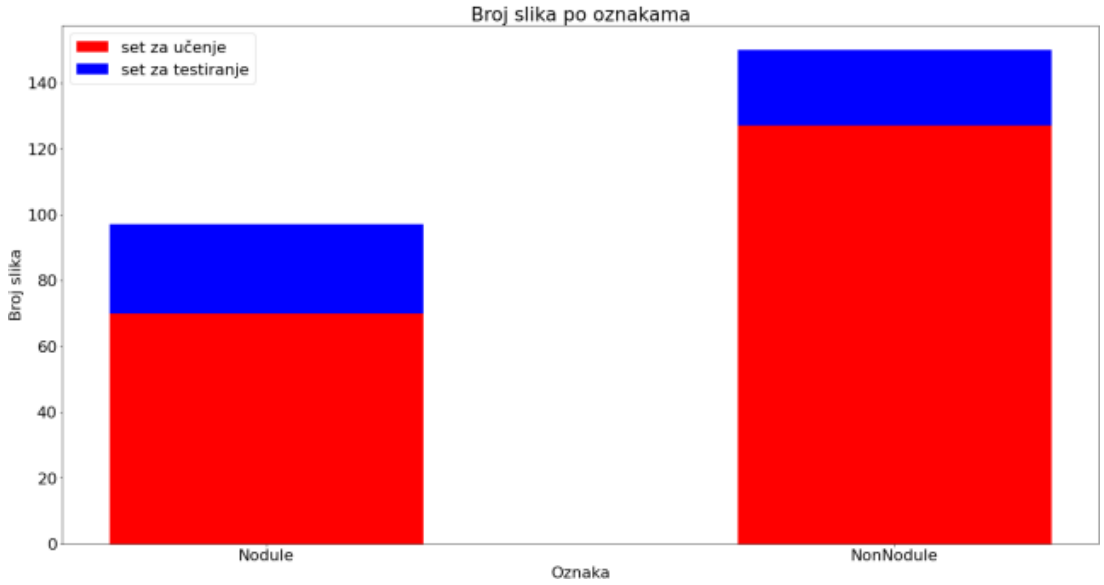
Izvor: Autori

Slika 3. *Raspodjela broja slika po oznakama za skup podataka Public Lung Dataset.*



Izvor: Autori

Slika 4. *Raspodjela broja slika po oznakama za skup podataka JSRT Dataset.*



Izvor: Autori

Arhitektura modela sastoji se od sljedećih 5 slojeva: sloj konvolucije, prvi kvadratni sloj, sloj sažimanja, drugi kvadratni sloj i izlazni sloj, prilagođenih za rad s homomorfno kriptiranim slikama. Usporedbom rezultata iz tablice 1 i tablice 2 može se zaključiti da predikcije na kriptiranim slikama pokazuju blago smanjenje učinkovitosti, no rezultati su zadovoljavajući. Na skupu ChestX-Det postignuta je mjera F1 od 0.9386 za nekriptirane slike i 0.8715 za kriptirane, što pokazuje dobru sposobnost klasifikacije unatoč kriptiranju. Slično, točnost na skupu Public Lung iznosi 84.22 % za nekriptirane i 80.83 % za kriptirane slike, dok točnost na JSRT skupu doseže 66 % za nekriptirane i 59.01 % za kriptirane slike. Blagi pad točnosti uočava se zbog kriptiranja, ali prednosti sigurnosti i zaštite privatnosti značajne su u kontekstu obrade osjetljivih medicinskih podataka.

Tablica 5. *Rezultati predikcije modela na slikama bez kriptiranja.*

	točnost (engl. *accuracy*)	F1 mjera (engl. F1 *score*)	preciznost (engl. *precistion*)	odziv (engl. *recall*)
ChestX-Det	-	0.9386	0.8843	1
Public Lung	84.22 %	-	-	-
JSRT	66 %	-	-	-

Izvor: Autori

Tablica 6. *Rezultati predikcije modela na kriptiranim slikama.*

	točnost (engl. *accuracy*)	F1-mjera (engl. F1-*score*)	preciznost (engl. *precistion*)	odziv (engl. *recall*)
ChestX-Det	-	0.8715	0.7431	1
Public Lung	80.83 %	-	-	-
JSRT	59.01 %	-	-	-

Izvor: Autori

5. Zaključak

U ovom radu analiziran je utjecaj primjene homomorfne enkripcije na učinkovitost neuronske mreže primjenom metode iz projekta CryptoNets. Rezultati pokazuju da homomorfna enkripcija nije bitno utjecala na točnost modela, iako je na svim skupovima podataka zabilježeno blago smanjenje točnosti koje je ostalo unutar zadovoljavajućih granica.

Na skupu ChestX-Det gdje je riječ o višeklasifikacijskom problemu, model je evaluiran pomoću F1 mjere, dok je na Public Lung Dataset i JSRT Dataset korištena mjera točnosti za evaluaciju klasifikacije. Rezultati su pokazali stabilnost modela čak i na kriptiranim podatcima pri čemu su F1 mjera i točnost ostali na zadovoljavajućim razinama.

Lošiji rezultati na JSRT skupu pripisani su ograničenom broju primjera u skupu što može otežati modelu učenje složenih obrazaca. Zaključeno je da modeli temeljeni na CryptoNets pružaju ravnotežu između sigurnosti i učinkovitosti, dok se budući radovi mogu usmjeriti na daljnju optimizaciju algoritama homomorfne kriptografije kako bi se poboljšali sigurnost podataka, učinkovitost modela i brzina izvođenja.

Literatura

Downlin, N. et al. (2016.). "CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy", *Proceedings of Machine Learning Research*, vol. 48, 20-22.

Krajčí, M. (2023.). "Utjecaj homomorfne kriptografije na sustav za prepoznavanje slika zasnovan na strojnom učenju", diplomski rad, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva

Kundan M., Bhatia R. (2023.). "A systematic review of homomorphic encryption and its contributions in healthcare industry", *Complex Intell*, vol. 9, 3759–3786
Lian J. et al. (2021.), "ChestX-Det-Dataset", https://github.com/Deepwise-AILab/ChestX-Det-Dataset (29.10.2024.)
Reeves, A. P. et al. (2009.), "A Public Image Database to Support Research in Computer Aided Diagnosis", *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3715-3718
Shiraishi J. et al. (2000.), "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules", *AJR*, vol. 174, 71-74
Žigrović, D. (2017.). "Homomorfna kriptografija u primjeni", završni rad, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva