

# Klasifikacija zdravlja fetusa na temelju kardiotokografije

## 1. Uvod

Fetalna patnja (*fetal distress*) je prilično slabo definiran medicinski pojam koji se odnosi na ono što se događa kad nerođeno dijete počinje imati problema, najčešće u trećem tromjesečju trudnoće. U većini slučajeva, fetalna patnja uključuje nedostatak protoka zraka za fetus što dovodi do izvođenja hitnog carskog reza. Kako bi se to spriječilo ili pravovremeno liječilo koristi se dijagnostička metoda praćenja stanja fetusa zvana kardiotokografija (CTG). Kardiotokografija podrazumijeva grafički prikaz aktivnosti srca ploda i aktivnosti mišića zida maternice tokom trudnoće i tokom samog porođaja.

U cilju očuvanja zdravlja fetusa, bitna je brza interpretacija CTG snimki, pa je CTG data set često korišten u radovima vezanim za strojno učenje i klasifikaciju. Većina dostupnih radova imaju u cilju podijeliti snimke u 2 klase: normalnu i patološku ili u 3 klase: normalnu (normal), sumnjivu (suspect) i patološku (pathogenic). Postoji i mogućnost raspodjele snimaka u 10 klasa koje precizno određuju stanje srca fetusa, ali raspodjela u te klase se rijetko koristi.

U svim prethodno objavljenim radovima koristi se Cardiotocography data set iz UCI repozitorija za strojno učenje [1]. Data set se sastoji od 2126 automatski procesiranih primjeraka CTG snimki. Svaka snimka je predstavljena pomoću 21 varijable (broj otkucaja u minuti, broj pomaka fetusa u minuti, broj kontrakcija maternice u minuti itd.). Sve snimke su klasificirane u prethodno navedene klase od strane stručnjaka.

## 2. Pregled literature

Diskriminantna analiza (DA), stabla odlučivanja (DT) i umjetne neuronske mreže (ANN) su korišteni za klasifikaciju danog dana seta u radu iz 2012. godine autora Huang. Postignuta je točnost od 82.1%, 86.36% i 97.78%, respektivno [2]. 2012. je također Sundar pomoću neuronskih mreža postigao točnost od 80% [3]. Yılmaz and Kılıkçier 2013. koriste metodu potpornih vektora (SVM) i dobivaju točnost 91.62% [4]. Ocak i Ertunç (2013.) klasificiraju podatke u 2 klase (normalnu i patološku) pomoću prilagodljivog neuro-neizrazitog sustava (ANFIS) i time točno klasificiraju 96.6% patoloških stanja i 97.2% normalnih stanja [5]. 2014. Karabulut uspoređuje 6 metoda strojnog učenja: naivni Bayes (NB), radijalne mreže (RBN), Bayesove mreže (BN), SVM, ANN i DT bez i s AdaBoost metodom. Točnost NB, RBN, BN i DT je 87.39%, 87.67%, 92.61% i 95.01% respektivno, a ostalim metodama nije postignut značajan napredak [6]. Korištenjem metode najbližih susjeda (k-NN) i slučajnih šuma (RF) Şahin i Subasi postižu točnost od 98.4% i 99.18% [7]. Performanse ANN i ELM (Extreme Learning Machine) su uspoređene u radu autora Cömert i dobivena je veća točnost pomoću ELM-a (93.42%) [8]. Arif predlaže upotrebu slučajnih šuma i dobiva točnost od 93.6% [9]. Kamath i Kamat primjenjuju istu metodu za podjelu u 10 klasa i dobivaju preciznost preko 87% [10].

### 3. Cilj projekta i plan rada

Cilj našeg projekta je pronalazak alata koji će biti koristan u detekciji mogućeg patološkog stanja. Eksploratornom analizom data seta dolazimo do zaključka da je dani problem nebalansiran klasifikacijski problem što autori spomenutih članaka zanemaruju. Uspješnost svojih modela najčešće opisuju pomoću točnosti (omjer točno klasificiranih instanci i ukupnog broja instanci), što nije idealno za nebalansirane klasifikacijske probleme.

Najveći problem kod klasifikacije nebalansiranih podataka je pogrešno pridruživanje instance iz manje zastupljene klase u zastupljeniju klasu, što bi u našem slučaju značilo pridruživanje kardiotokograma iz klase suspect ili pathogenic klasi normal. To svakako želimo izbjeći, jer je za zdravlje djeteta važno prepoznati abnormalan kardiotokogram.

Za evaluaciju modela ćemo stoga uz točnost koristiti i konfuzijsku matricu, osjetljivost i preciznost.

U rješavanju problema nebalansiranosti ćemo uz Python biblioteku scikit-learn koristiti i imbalanced-learn.

Dakle, u našem projektu naglasak će biti na pokušaju rješavanja problema nebalansiranosti, stoga ćemo isprobati nekoliko različitih metoda strojnog učenja i međusobno ih usporediti.

### 4. Očekivani rezultati

Budući da cilj ovog projekta nije postizanje najveće točnosti, već proučavanje problema nebalansiranosti podataka, očekujemo da točnost našeg rješenja neće biti u rangu s trenutno najboljim ponuđenim rješenjima (točnost >99%), ali vjerujemo da će i naš model davati zadovoljavajuće rezultate.

### 5. Literatura

[1] <https://archive.ics.uci.edu/ml/datasets/cardiotocography>

[2] M.-L. Huang and Y.-Y. Hsu, "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network," *Journal of Biomedical Science and Engineering*, vol. 5, p. 526, 2012.

[3] C. Sundar, et al., "Classification of cardiotocogram data using neural network based machine learning technique," *International Journal of Computer Applications*, vol. 47, 2012.

[4] E. Yılmaz and Ç. Kılıkçier, "Determination of fetal state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree," *Computational and mathematical methods in medicine*, vol. 2013, 2013.

[5] H. Ocak and H. M. Ertunc, "Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems," *Neural Computing and Applications*, vol. 23, pp. 1583-1589, 2013.

[6] E. M. Karabulut and T. Ibrikci, "Analysis of cardiotocogram data for fetal distress determination by decision tree based adaptive boosting approach," *Journal of Computer and Communications*, vol. 2, p. 32, 2014.

- [7] H. Sahin and A. Subasi, "Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques," *Applied Soft Computing*, vol. 33, pp. 231-238, 2015.
- [8] Z. Cömert, et al., "Cardiotocography signals with artificial neural network and extreme learning machine," in *Signal Processing and Communication Application Conference (SIU)*, 2016 24th, 2016, pp. 1493-1496.
- [9] M. Arif, "Classification of cardiotocograms using random forest classifier and selection of important features from cardiotocogram signal," *Biomaterials and Biomechanics in Bioengineering*, vol. 2, pp. 173-183, 2015.
- [10] R. Kamath and R. Kamat, "Modeling fetal morphologic patterns through cardiotocography data: Decision tree-based approach," *Journal of Pharmacy Research| Vol*, vol. 12, p. 10, 2018.