KO'KRAK SARATONI KASALLIGINI TASHHISLASHDA MASHINALI O'RGATISH MODELLARINING QIYOSIY TAHLILI

Tursunmurotov Davrbek Xudayorovich[1], Normamatov Xayriddin Mengniyevich[2], Karimova Dilfiya Shuxratovna[2]

[1]Axborot texnologiyalari va menejment universiteti
[2]Osiyo texnologiyalari universiteti
E-mail:mr.davrbek@mail.ru

K E Y W O R D S

mashinali o'rgatish, ko'krak saratoni, erta tashhislash, mashinali o'rgatish, sun'iy intellekt, tasniflash algoritmlari

A B S T R A C T

Ushbu maqolada ko'krak saratonini erta tashhislash masalasi mashinali o'rganish algoritmlaridan foydalanish asosida tadqiq etiladi. Tadqiqotning asosiy maqsadi ko'krak bezidagi o'smalarning xavfli (malignant) yoki xavfsiz (benign) ekanligini hujayra yadrolarining morfologik xususiyatlarini ifodalovchi miqdoriy belgilar yordamida aniqlash imkoniyatlarini baholashdan iborat[2]. Tadqiqot uchun Breast Cancer Wisconsin (Diagnostic) ma'lumotlar to'plamidan foydalanildi. Ushbu tanlanmada har bir obyekt hujayra radiusi, perimetri, maydoni, silliqligi, simmetriyasi va fraktal o'lchami kabi ko'rsatkichlar bilan tavsiflanadi.

Eksperiment jarayonida Logistic Regression[5], K-Nearest Neighbors[4] va Support Vector Machine (RBF yadrosi asosida)[3] algoritmlari qo'llanilib, ularning tasniflash samaradorligi qiyosiy tahlil qilindi. Ma'lumotlar oldindan normallashtirilib, modellar bir xil sharoitda o'rgatildi hamda test tanlanmada sinovdan o'tkazildi. Modellarning samaradorligi aniqlik (accuracy), aniqlik koeffitsienti (precision), to'liq qamrov (recall) va F1-o'lchovi mezonlari asosida baholandi.

Olingan natijalar mashinali o'rgatish modellarining yuqori aniqlik bilan ishlashini hamda ko'krak saratonini tashhislash jarayonida tibbiy mutaxassislar uchun qaror qabul qilishni qo'llab-quvvatlovchi samarali vosita bo'lishi mumkinligini ko'rsatdi. Tadqiqot sun'iy intellekt texnologiyalarining diagnostika jarayonlarini takomillashtirishdagi amaliy ahamiyatini tasdiqlaydi.

KIRISH

Hozirgi kunda ko'krak saratoni ayollar orasida eng keng tarqalgan va eng xavfli onkologik kasalliklardan biri hisoblanadi. Jahon sog'liqni saqlash tashkiloti ma'lumotlariga ko'ra, har yili millionlab ayollarda ushbu kasallik aniqlanib, ularning katta qismi aynan kech tashhis qo'yilishi oqibatida hayotdan ko'z yumadi. Ko'krak saratonining dastlabki bosqichlarida aniqlanishi bemorlarning yashab qolish ehtimolini sezilarli darajada oshiradi hamda davolash jarayonini samaraliroq va kam xarajatli qiladi. Shu sababli ushbu kasallikni erta tashhislash masalasi zamonaviy tibbiyotning eng dolzarb muammolaridan biri bo'lib qolmoqda.

An'anaviy tashxis qo'yish usullari, jumladan, biopsiya, mammografiya va shifokorlarning vizual tahliliga asoslangan qarorlar yuqori malaka va katta tajribani talab qiladi. Biroq inson omilini ta'siri, vaqt cheklovlari va tashxisdagi xatoliklar ayrim holatlarda noto'g'ri qarorlar qabul qilinishiga olib kelishi mumkin. Ushbu muammolarni kamaytirish va tashxis jarayonini yanada aniqlashtirish maqsadida zamonaviy axborot texnologiyalari, xususan, sun'iy intellekt va mashinali o'rgatish algoritmlaridan foydalanish tobora muhim ahamiyat kasb etmoqda.

Ko'krak saratonini aniqlash jarayoni odatda bir nechta klinik va laboratoriya tekshiruv bosqichlaridan iborat bo'lib, ular o'smalarning tabiati va rivojlanish darajasini aniqlashga qaratilgan. Dastlabki bosqichlarda klinik ko'rik va tasviriy tekshiruvlar amalga oshirilsa, yakuniy tashxis ko'pincha biopsiya asosida olingan hujayra

© D.X. Tursunmurotov, X.M. Normamatov, D.Sh. Karimova

namunalarining tahliliga tayanadi. Xususan, Fine Needle Aspiration (FNA) usuli yordamida olingan hujayra yadrolarining shakli, o‘lchami va tuzilish xususiyatlari o‘smalarning xavfli yoki xavfsiz turga mansubligini aniqlashda muhim ahamiyatga ega. FNA – bu o‘ta ingichka igna yordamida shubhali o‘sma yoki to‘qimadan mayda hujayra namunalarini so‘rib olish usulidir.

Mazkur tadqiqotda qo‘llanilgan Breast Cancer ma’lumotlar to‘plami[8] aynan FNA asosida olingan hujayra yadrolarining raqamli qiymatlaridan iborat miqdoriy alomatlarni o‘z ichiga oladi. Ushbu alomatlar hujayra yadrosining radiusi, perimetri, maydoni, silliqligi, notekisligi va simmetriyasi kabi morfologik belgilarni ifodalaydi. Amaliy tashxislashda ushbu alomatlarni vizual baholash mutaxassis tajribasiga bog‘liq bo‘lib, ayrim chegaraviy holatlarda subyektiv xatoliklarga olib kelishi mumkin.

So‘nggi yillarda mashinali o‘rgatish algoritmlari katta hajmdagi tibbiy ma’lumotlarni tahlil qilish, yashirin qonuniyatlarni aniqlash va yuqori aniqlikdagi bashoratlar berish imkoniyati bilan tibbiyot sohasida keng qo‘llanila boshladi. Ayniqsa, ko‘krak saratoni kabi murakkab kasalliklarni tashhislashda mashinali o‘rgatish modellari shifokorlar uchun qaror qabul qilish jarayonida samarali yordamchi vosita bo‘lib xizmat qilishi mumkin. Ushbu yondashuvlar inson omiliga bog‘liq xatoliklarni kamaytirish, tashxis qo‘yish tezligini oshirish va kasallikni erta bosqichlarda aniqlash imkonini beradi.

Maqolaning asosiy maqsadi sun’iy intellekt algoritmlarining ko‘krak saratonini erta tashxislashdagi samaradorligini tahlil qilish hamda ularning amaliy ahamiyatini ko‘rsatib berishdan iborat.

Sun’iy intellektni tibbiyotda qo‘llanishi

Tibbiyot sohasida raqamli texnologiyalar rivojlanishi natijasida katta hajmdagi klinik va laboratoriya ma’lumotlar shakllandi. Ushbu ma’lumotlarni an’anaviy usullar bilan tahlil qilish murakkablashib, sun’iy intellekt texnologiyalarini qo‘llash zaruratini yuzaga keltirdi. Shu bilan birga, sun’iy intellekt tizimlari mustaqil tashxis qo‘yuvchi mexanizm sifatida emas, balki shifokorlarning tajribasini to‘ldiruvchi va xatolik ehtimolini kamaytiruvchi yordamchi vosita sifatida qaraladi. Ushbu yondashuv tibbiyotda sun’iy intellekt texnologiyalarining istiqbolli va barqaror rivojlanishiga asos yaratmoqda. Shu sababli sun’iy intellekt algoritmlari, kasalliklarni tashxislashda keng qo‘llanila boshladi. Hozirgi kunda sun’iy intellekt tibbiyotda asosan qaror qabul qilishni qo‘llab-quvvatlovchi vosita sifatida ishlatilmoqda. U tibbiy tasvirlarni avtomatik tahlil qilish, kasalliklarni erta bosqichlarda aniqlash va tashxis aniqligini oshirish imkonini beradi.

Sun’iy intellektning tibbiyot sohasida qo‘llanilishi bir necha tarixiy bosqichlar orqali shakllangan bo‘lib, har bir bosqich o‘ziga xos yondashuvlar va texnologik imkoniyatlar bilan tavsiflanadi.

Ekspert tizimlar. Ushbu bosqichda tibbiyotda sun’iy intellekt asosan ekspert tizimlar ko‘rinishida qo‘llanilgan. MYCIN kabi tizimlar shifokorlarning bilimlarini “agar–unda” (IF–THEN) qoidalari shaklida modellashtirib, qon infeksiyalariga tashxis qo‘yishda foydalanilgan. Bu yondashuv murakkab klinik qarorlarni avtomatlashtirish imkonini bergan bo‘lsa-da, tizimlar yangi ma’lumotlardan mustaqil o‘rganish qobiliyatiga ega emas edi. Natijada, ularning moslashuvchanligi past bo‘lib, zamonaviy tibbiy sharoitlarga tez moslasha olmagan.

Statistik va mashinali o’rgatish yondashuvlari. Ushbu davrda tibbiyotda sun’iy intellekt qoidaga asoslangan tizimlardan statistik va ma’lumotlarga tayangan machine learning modellarga o‘ta boshladi. Logistic Regression[5], Decision Tree, Support Vector Machine (SVM)[2] va K-Nearest Neighbors (KNN)[3] kabi algoritmlar klinik ma’lumotlar asosida tashxis qo‘yishda keng qo‘llanildi. Aynan shu davrda yaratilgan *Breast Cancer* berilganlar to‘plami tibbiy klassifikatsiya masalalarida ko’krak saratoniga tashxislashdagi muhim berilganlardan biri bo’lib kelmoqda. Mazkur dataset mashinali o‘rgatish algoritmlarining saraton kasalligini erta aniqlashdagi samaradorligini baholashda keng foydalanilgan. Bu bosqich

© D.X. Tursunmurotov, X.M. Normamatov, D.Sh. Karimova

tibbiyotda ma’lumotlarga asoslangan tashxis qo‘yish konsepsiyasining shakllanishiga asos bo‘ldi.

Deep Learning (hozirgi davr). Ushbu bosqich tibbiyotda sun’iy intellektning jadal rivojlanishi bilan tavsiflanadi va chuqur neyron tarmoqlarga asoslangan yondashuvlar yetakchi o‘rinni egalladi. Ayniqsa, Convolutional Neural Networks (CNN) tibbiy tasvirlarni tahlil qilishda yuqori aniqlik ko‘rsatkichlarini ta’minlab, mammografiya, MRT va KT kabi medical imaging ma’lumotlarida keng qo‘llanilmoqda. Big Data texnologiyalari katta hajmdagi klinik va tasviriy ma’lumotlarni qayta ishlash imkonini berdi. Shuningdek, multimodal tizimlar tasvir, klinik ko‘rsatkichlar va genetik ma’lumotlarni birgalikda tahlil qilib, tashxis aniqligini yanada oshirdi. Natijada, ko‘krak bezi saratonini erta aniqlash va bashoratlashda deep learning asosidagi tizimlar muhim ahamiyat kasb etmoqda.

Tadqiqotda foydalanilgan tanlanma

Sun’iy intellektga asoslangan tashxislash tizimlarining muvaffaqiyati nafaqat tanlangan algoritmlarga, balki ular foydalanadigan tanlanmaning sifati va ishonchliligiga ham bevosita bog‘liq. Yuqorida ko‘rib chiqilgan sun’iy intellekt yondashuvlarini amaliy jihatdan baholash uchun real tibbiy ma’lumotlarga asoslanish zarur. Shu maqsadda ushbu tadqiqotda ko‘krak saratoni tashxisiga oid ochiq berilganlardan foydalanildi.

Mazkur tadqiqotda tanlanma sifatida **Breast Cancer Wisconsin (Diagnostic)[8]** tanlanmasi olindi. Ushbu tanlanma AQShning Wisconsin universitetida olib borilgan klinik tadqiqotlar natijasida shakllantirilgan bo‘lib, ko‘krak bezidagi o‘smalardan olingan hujayra namunalarini sitologik tahlil qilish asosida tuzilgan. Ma’lumotlar **Fine Needle Aspiration (FNA)** usuli yordamida olingan bo‘lib, bu usul o‘smadan ingichka igna orqali hujayra namunasi olishga asoslanadi va tibbiyot amaliyotida keng qo‘llaniladi.

Tanlanma jami 569 ta obyektdan iborat bo‘lib, har bir obyekt 32 ta alomatdan iborat alohida bemorning holatini ifodalaydi. Tashxis natijasi ikki sinfdan iborat: *malign (M)* — yomon sifatli, ya’ni saratonli o‘sma va *benign (B)* — yaxshi sifatli, ya’ni saratonsiz o‘sma. Alomatlar hujayra radiusi, perimetri, maydoni, teksturasi, silliqligi, simmetriyasi va fraktal o‘lcham kabi ko‘rsatkichlarni o‘z ichiga oladi. Har bir xususiyat uch xil ko‘rinishda berilgan: o‘rtacha qiymat (mean), standart xatolik (standard error) va eng yomon holat (worst value). Bunday tuzilma hujayra shaklining nafaqat umumiy holatini, balki uning o‘zgaruvchanligi va ekstremal qiymatlarini ham hisobga olish imkonini beradi.

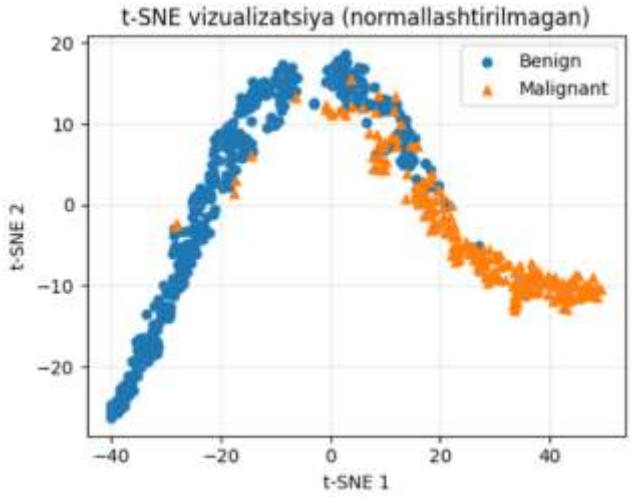
Tanlanmada yo‘qolgan qiymatlar mavjud emas, ya’ni u to‘liq va oldindan qayta ishlashni talab qilmaydi. Bu esa mashinali o‘rgatish algoritmlarini qo‘llash jarayonini soddalashtiradi va natijalarning ishonchliligini oshiradi. Mazkur tanlanma ko‘plab ilmiy tadqiqotlarda benchmark sifatida foydalanilib kelinmoqda va ko‘krak saratoni kasalligini erta tashhislash masalasida sun’iy intellekt algoritmlarining samaradorligini baholash uchun qulay hisoblanadi.

Yuqorida tavsiflangan ma’lumotlar to‘plamining fazoviy tuzilmasini vizual jihatdan baholash hamda sinflar o‘rtasidagi ajralish darajasini aniqlash maqsadida qo‘shimcha vizual tahlil amalga oshirildi. Buning uchun ko‘p o‘lchamli belgilar fazosini past o‘lchamli fazoga proyeksiyalovchi t-Distributed Stochastic Neighbor Embedding (t-SNE)[6] usulidan foydalanildi. Mazkur usul nochiziqli o‘lcham kamaytirish algoritmlaridan biri bo‘lib, yuqori o‘lchamli ma’lumotlar orasidagi lokal o‘xshashliklarni saqlagan holda ularni ikki yoki uch o‘lchamli fazoda tasvirlash imkonini beradi. Shu sababli t-SNE ma’lumotlar klasterlashuvi hamda sinflar ajralishini vizual baholashda keng qo‘llaniladi.

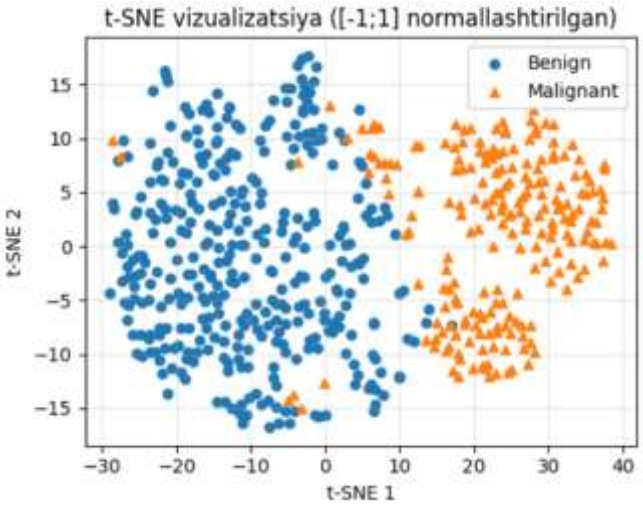
Dastlab ma’lumotlar dastlabki holatda, ya’ni belgilar diapazonlari normallashtirilmasdan turib vizualizatsiya qilindi (1-rasm). Mazkur holatda belgilar o‘lcham diapazonlari o‘rtasidagi sezilarli tafovutlar sababli fazoviy tasvirda obyektlarning tartibli klasterlashuvi sust kuzatildi. Xususan, benign va malignant sinflarga mansub obyektlar umumiy fazoda qisman aralashgan holda joylashgani kuzatildi. Bu holat ayrim belgilar

© D.X. Tursunmurotov, X.M. Normamatov, D.Sh. Karimova

diapazonining kattaligi masofa hisoblash jarayoniga dominant ta'sir ko'rsatishi bilan izohlanadi.



1-rasm. Breast Cancer tanlanmasini t-SNE da tasvirlanishi



1-rasm. [-1;1] da normallsahtirilgan Breast Cancer tanlanmasini t-SNE da tasvirlanishi

Keyingi bosqichda belgilar diapazonlarini yagona masshtabga keltirish maqsadida ma'lumotlar [-1;1] oraliqda normallashtirildi va t-SNE algoritmi qayta qo'llandi (2-rasm). Normallashtirish natijasida barcha alomatlar masofa hisoblash jarayonida teng ulush bilan ishtirok eta boshladi. Natijada tasvirda sinflar ajralishi ancha yaqqol namoyon bo'ldi: benign obyektlar nisbatan zich klaster hosil qilgan bo'lsa, malignant obyektlar alohida fazoviy sohada joylashgani kuzatildi.

Olingan vizual natijalar ko'krak saratoni tashhislash masalasida tanlangan belgilar to'plami tasniflash algoritmlarini qo'llash uchun yetarli darajada informativ ekanligini tasdiqlaydi. Shu bilan birga, alomatlarni normallashtirish jarayoni nafaqat mashinali o'rganish modellarini o'qitishda, balki obyektlarning tuzilmasini vizual aks ettirishda muhim ahamiyatga ega.

Qo'llanilgan sun'iy intellekt algoritmlari

Yuqorida tavsiflangan tanlanma ko'krak saratoni kasalligini tashhislash masalasini sun'iy intellekt yondashuvi asosida o'rganish uchun yetarli statistik va morfologik ma'lumotlarni o'z ichiga oladi. Mazkur tanlanma asosida tasniflash modellarini qurish va ularning samaradorligini baholash uchun turli mashinali o'rganish algoritmlaridan foydalanish maqsadga muvofiq hisoblanadi. Sun'iy intellekt algoritmlari tasniflash masalalarini hal etishda turli nazariy va matematik yondashuvlarga asoslanadi. Har bir algoritm ma'lumotlarni tahlil qilish, sinflar orasidagi chegarani aniqlash hamda yangi obyektlarni klassifikatsiya qilish jarayonini o'ziga xos matematik model asosida amalga oshiradi. Ayrim algoritmlar ehtimollik nazariyasiga tayansa, boshqalari masofaga asoslangan o'xshashlik mezonlari yoki optimal ajratuvchi chegara qurish prinsiplari asosida ishlaydi. Shu bois turli algoritmlarni bir xil ma'lumotlar to'plamida qo'llash ularning tasniflash samaradorligini har xil nuqtai nazardan baholash imkonini beradi. Mazkur tadqiqotda ko'krak saratonini tashhislash masalasini yechish uchun turli matematik yondashuvlarga asoslangan Logistic Regression, K-Nearest Neighbors va Support Vector Machine algoritmlaridan foydalanildi[2,3,4].

Logistik Regressiya ikkilik tasniflash masalalarini hal etishda keng qo'llaniladigan statistik mashinali o'rganish algoritmlaridan biri hisoblanadi. Mazkur algoritm kirish belgilarining chiziqli kombinatsiyasi asosida obyektning ma'lum sinfga tegishliligini ehtimollik asosida baholaydi. Ko'krak saratonini tashhislash masalasida Logistik Regressiya modeli hujayra morfologik xususiyatlari asosida o'smalarning xavfli yoki xavfsiz turga mansubligini ehtimollik nuqtai nazaridan aniqlash imkonini beradi.

© D.X. Tursunmurotov, X.M. Normamatov, D.Sh. Karimova

Algoritmning matematik asosi logistika (sigmoid) funksiyasiga tayanadi:

$$P(y=1|x)=\frac{1}{1+e^{-z}} \qquad (1)$$

bu yerda

$$z = w^T x + b$$

x – alomatlar
w – vazn koeffitsientlari
b – ozod had

Model chiqishida 0 va 1 oralig'idagi ehtimollik qiymati olinadi. Ushbu qiymat berilgan chegaraviy (threshold) qiymat bilan solishtirilib, obyekt ***benign*** yoki ***malignant*** sinfiga ajratiladi. Logistic Regression modelining afzalligi uning soddaligi, interpretatsiya qilish qulayligi hamda tibbiy qarorlarni tushuntirish imkoniyati yuqoriligidadir.

K-Nearest Neighbors algoritmi masofaga asoslangan tasniflash usullaridan biri bo'lib, u yangi obyektni o'quv tanlanmasidagi eng yaqin qo'shni obyektlar sinfiga qarab klassifikatsiya qiladi. Ushbu algoritmda oldindan model qurilmaydi; tasniflash butunlay alomatlar fazosidagi o'xshashlik darajasiga asoslanadi. Ko'krak saratoni tashxislashda bu yondashuv o'xshash hujayra tuzilishiga ega namunalarni bir sinfga kiritish imkonini beradi.

Ikki obyekt orasidagi masofa odatda Evklid metrikasi orqali aniqlanadi:

$$d(x_i, x_j)=\sqrt{\sum_{k=1}^{n}(x_{ik}-x_{jk})^2} \qquad (2)$$

bu yerda n – alomatlar soni

KNN algoritmi berilganlarning lokal soxa tuzilishini hisobga olishi sababli tibbiy diagnostika masalalarida, ayniqsa ayniqsa obyektlar o'xshashligini aniqlashda samarali natijalar beradi.

Support Vector Machine algoritmi tasniflash masalalarini hal etishda optimal ajratuvchi chegara qurishga asoslangan kuchli mashinali o'rganish usullaridan biridir. Algoritmning asosiy maqsadi turli sinf obyektlarini maksimal masofa (margin) bilan ajratadigan gipertekislikni aniqlashdan iborat. Bu esa modelning umumlashtirish qobiliyatini oshiradi.

Chiziqli SVM modelining matematik ifodasi quyidagicha beriladi:

$$w^T x + b = 0$$

bu yerda ajratuvchi tekislik sinflar orasidagi masofa maksimal bo'lishi sharti bilan aniqlanadi. Optimallashtirish masalasi quyidagi ko'rinishda yozilib, marginni maksimal qilishga yo'naltiriladi:

$$\min \frac{1}{2}\|w\|^2$$

Nochiziqli tasniflash holatlarida esa kernel funksiyalar qo'llanilib, ma'lumotlar yuqori o'lchamli fazoga akslantiriladi.

SVM algoritmi yuqori o'lchamli ma'lumotlar bilan samarali ishlashi, sinflar chegarasini aniq ajratishi hamda yuqori aniqlik ko'rsatkichlariga ega bo'lishi sababli tibbiy diagnostika masalalarida, xususan, ko'krak saratonini aniqlashda keng qo'llaniladi[2].

Yuqorida keltirilgan algoritmlar asosida ko'krak saratoni tashxislash masalasini amaliy jihatdan baholash maqsadida hisoblash eksperimenti o'tkazildi. Eksperiment natijalari va ularning tahlili keyingi bo'limda keltiriladi.

Hisoblash eksperimenti

Mashinali o'rganish algoritmlarining nazariy jihatlari ko'rib chiqilgach, ularning ko'krak saratonini tashhislash masalasidagi amaliy imkoniyatlarini aniqlash maqsadida hisoblash eksperimenti o'tkazildi. Eksperiment davomida tanlangan tasniflash modellarining barchasi bir xil tanlanma asosida o'qitildi hamda test sinovlaridan o'tkazildi. Olingan natijalar tasniflash samaradorligini ifodalovchi qator baholash mezonlari asosida qiyosiy tahlil qilindi.

© D.X. Tursunmurotov, X.M. Normamatov, D.Sh. Karimova

Eksperiment Python dasturlash muhiti va scikit-learn kutubxonasi yordamida amalga oshirildi[9]. Modellarni o‘rgatish va sinovdan o‘tkazish jarayonida ma’lumotlar o‘rgatuvchi va test to‘plamlariga ajratildi. Tanlanma obyektlari tasodifiy aralashtirilgan holda 80% o‘quv va 20% test nisbatida bo‘lindi. Bunday ajratish modellarni o‘rgatish va ularning umumlashtirish qobiliyatini baholash imkonini beradi.

Model barqarorligini oshirish hamda alomatlar diapazonlari o‘rtasidagi tafovutni bartaraf etish maqsadida ma’lumotlar normallashtirildi. Normallashtirilgan ma’lumotlar asosida Logistic Regression, K-Nearest Neighbors va Support Vector Machine algoritmlari alohida o‘qitildi[2,3,4].

Modellar samaradorligini baholash uchun bir nechta statistik mezonlardan foydalanildi. Xususan, tasniflash aniqligi Accuracy, Precision, Recall hamda F1-o‘lchovi hisoblandi va modellar qiyosiy tahlil qilindi.

1-jadval.

Tasniflash modellarining samaradorlik ko‘rsatkichlari bo‘yicha qiyosiy tahlili

Model	Accuracy	Precision	Recall	F1-score
Logistik regressiya	0.982456	1	0.9523	0.9756
KNN(k=5)	0.9642	1	0.9047	0.9500
SVM(RBF)	0.9824	1	0.9523	0.9756

1-jadvalda keltirilgan natijalardan ko‘rinib turibdiki, Logistic Regression va SVM (RBF) modellarining samaradorlik ko‘rsatkichlari deyarli bir xil qiymatga ega bo‘ldi. Bu holat, avvalo, tanlangan belgilar to‘plamining yuqori informativligi hamda sinflar o‘rtasidagi ajralishning nisbatan aniq ekanligi bilan izohlanadi. Shuningdek, ma’lumotlarning [-1;1] diapazonda normallashtirilishi barcha belgilarni yagona masshtabga keltirib, masofaga asoslangan hisob-kitoblarda ularning teng ulush bilan ishtirok etishini ta’minladi. Bu esa ayniqsa SVM va KNN kabi masofaga sezgir modellar uchun tasniflash aniqligini oshirishga xizmat qildi.

Natijalarga ko‘ra, ko‘krak saratoni tashhislash masalasida chiziqli model (Logistic Regression) ham yuqori aniqlik bilan ishlay olishi kuzatildi. Bu mazkur tanlanmada sinflar o‘rtasidagi ajralish asosan chiziqli xarakterga ega ekanligini ko‘rsatadi. SVM (RBF) modeli nolinear ajralishni hisobga olish imkoniyatiga ega bo‘lsa-da, ushbu tanlanmada uning ustunligi sezilarli darajada namoyon bo‘lmadi, bu esa belgilar fazosining murakkab nolinear chegaralarni talab qilmasligini anglatadi.

KNN (k=5) algoritmi boshqa modellarga nisbatan biroz pastroq natija ko‘rsatdi. Buni KNN algoritmining lokal qo‘shnilar asosida ishlashi hamda parametr (k) tanloviga sezgirligi bilan izohlash mumkin. Shuningdek, KNN masofaviy hisoblashga to‘liq tayanishi sababli, hatto normallashtirish qo‘llanilgan taqdirda ham, chegaraga yaqin joylashgan obyektlar tasnifida xatolik ehtimoli yuqoriroq bo‘lishi mumkin.

Jadvaldan yana bir muhim jihat ko‘zga tashlanadi: barcha modellar uchun precision ko‘rsatkichi 1 ga teng bo‘ldi. Bu model tomonidan “malignant” deb tasniflangan barcha holatlar haqiqatan ham saratonli ekanligini, ya’ni noto‘g‘ri ijobiy (False Positive) xatolar kuzatilmaganligini anglatadi. Klinik nuqtai nazardan bu juda muhim natija hisoblanadi, chunki sog‘lom bemorga noto‘g‘ri tashxis qo‘yilishining oldi olinadi. Shu bilan birga, recall ko‘rsatkichining 1 dan kichik bo‘lishi ayrim saraton holatlari aniqlanmay qolganligini ko‘rsatadi, bu esa modelni yanada takomillashtirish zarurligini bildiradi.

Umuman olganda, olingan natijalar normallashtirilgan belgilar asosida mashinali o‘rgatish algoritmlarining ko‘krak saratonini tashhislash masalasida yuqori samaradorlik ko‘rsatishini tasdiqlaydi.

FOYDALANILGAN ADABIYOTLAR

1. **Dua D., Graff C.** UCI Machine Learning Repository. – Irvine, CA: University of

© D.X. Tursunmurotov, X.M. Normamatov, D.Sh. Karimova

California, School of Information and Computer Science, 2019.
2. **Wolberg W. H., Street W. N., Mangasarian O. L.** Machine learning techniques to diagnose breast cancer from fine-needle aspirates // Cancer Letters. – 1994. – Vol. 77. – pp. 163–171.
3. **Cortes C., Vapnik V.** Support-vector networks // Machine Learning. – 1995. – Vol. 20. – pp. 273–297.
4. **Cover T., Hart P.** Nearest neighbor pattern classification // IEEE Transactions on Information Theory. – 1967. – Vol. 13. – pp. 21–27.
5. **Hosmer D. W., Lemeshow S., Sturdivant R. X.** Applied Logistic Regression. – 3rd ed. – New York: Wiley, 2013.
6. **Van der Maaten L., Hinton G.** Visualizing data using t-SNE // Journal of Machine Learning Research. – 2008. – Vol. 9. – pp. 2579–2605.
7. **Goodfellow I., Bengio Y., Courville A.** Deep Learning. – Cambridge: MIT Press, 2016.
8. Kaggle – Breast Cancer Dataset: https://www.kaggle.com/datasets
9. Scikit-learn Documentation:https://scikit-learn.org

© *D.X. Tursunmurotov, X.M. Normamatov, D.Sh. Karimova*