SUN'IY INTELLEKT YORDAMIDA DIABET KASALLIGINI ERTA ANIQLASH MODELINI ISHLAB CHIQISH

Otaqulov Oybek Xamdamovich,
Farg'ona davlat texnika universiteti
Farg'ona, O'zbekiston
E-mail: oybek.otagulov.63@bk.ru

Nabiyev Iskandar Farxodjon o'g'li,
Farg'ona davlat texnika universiteti
Farg'ona, O'zbekiston
E-mail: inabiyev088@gmail.com

Annotatsiya. Ushbu maqolada diabet kasalligini erta aniqlash uchun sun'iy intellektga asoslangan model ishlab chiqish masalasi ko'rib chiqiladi. Tadqiqotda Pima Indians Diabetes Database ma'lumotlar to'plamidan foydalanilgan. Model turli klinik parametrlar asosida bemorda diabet mavjudligini bashorat qiladi. Olingan natijalar sun'iy intellekt metodlari tibbiyotda samarali qo'llanishi mumkinligini ko'rsatdi.

Kalit so'zlar: sun'iy intellekt, diabet, klassifikatsiya, machine learning, tibbiy diagnostika, Pima dataset

Kirish

Diabetes mellitus zamonaviy tibbiyotning eng dolzarb muammolaridan biri bo'lib, u metabolik buzilishlar guruhi sifatida tavsiflanadi va qonda glyukoza miqdorining uzoq muddat davomida yuqori darajada saqlanishi bilan xarakterlanadi. Ushbu kasallik inson organizmidagi insulin ishlab chiqarilishi yoki uning samarali ishlatilishidagi muammolar bilan bog'liq bo'lib, vaqt o'tishi bilan yurak-qon tomir tizimi, buyraklar, ko'z va asab tizimiga jiddiy zarar yetkazishi mumkin. Jahon miqyosida diabet kasalligi tez sur'atlarda ortib borayotgan global sog'liq muammosi hisoblanadi va u nafaqat tibbiy, balki ijtimoiy-iqtisodiy jihatdan ham katta xavf tug'dirmoqda.

Diabetning asosiy ikki turi mavjud bo‘lib, 1-tip diabet insulin yetishmovchiligi natijasida yuzaga keladi, 2-tip diabet esa asosan insulin rezistentligi bilan bog‘liq. Ayniqsa, 2-tip diabet keng tarqalgan bo‘lib, umumiy kasallanish holatlarining katta qismini tashkil etadi. Ushbu kasallikning yashirin rivojlanishi va dastlabki bosqichlarda aniq simptomlar bermasligi uni erta aniqlashni murakkablashtiradi. Shu sababli, kasallikni erta bosqichda aniqlash va oldini olish zamonaviy tibbiyotning ustuvor yo‘nalishlaridan biri hisoblanadi.

So‘nggi yillarda axborot-kommunikatsiya texnologiyalari va sun’iy intellekt metodlarining jadal rivojlanishi tibbiy diagnostika jarayonlarini sezilarli darajada takomillashtirish imkonini berdi. Xususan, mashinali o‘rganish algoritmlari katta hajmdagi tibbiy ma’lumotlarni tahlil qilish, yashirin qonuniyatlarni aniqlash va kasalliklarni bashorat qilishda yuqori samaradorlikni namoyon etmoqda. Bu esa an’anaviy diagnostika usullariga nisbatan tezkorlik, aniqlik va avtomatlashtirish darajasini oshirishga xizmat qilmoqda.

Mazkur tadqiqotda diabet kasalligini erta aniqlash masalasi sun’iy intellekt yondashuvlari asosida ko‘rib chiqiladi. Tadqiqot uchun Pima Indians Diabetes Database ma’lumotlar to‘plamidan foydalanildi. Ushbu ma’lumotlar to‘plami turli klinik va biologik ko‘rsatkichlarni o‘z ichiga olgan bo‘lib, ular yordamida bemorda diabet mavjudligini aniqlash imkoniyati mavjud. Jumladan, homiladorlik soni, glyukoza darajasi, qon bosimi, tana massasi indeksi, insulin miqdori va boshqa parametrlar asosiy omillar sifatida qaraladi.

Tadqiqotning asosiy maqsadi mavjud klinik parametrlar asosida diabet kasalligini bashorat qiluvchi samarali modelni ishlab chiqish va uning aniqlik darajasini baholashdan iborat. Shu bilan birga, turli mashinali o‘rganish algoritmlarining samaradorligi o‘zaro solishtirilib, eng optimal yechim aniqlanadi. Ushbu yondashuv nafaqat diagnostika jarayonini soddalashtirish, balki kasallikni erta aniqlash orqali uning asoratlarini kamaytirishga ham xizmat qiladi.

Natijada, ushbu tadqiqot sun’iy intellekt texnologiyalarining tibbiyot sohasidagi amaliy ahamiyatini ko‘rsatib berishga va diabet kabi keng tarqalgan kasalliklarni aniqlashda yangi innovatsion yechimlarni taklif etishga qaratilgan.

Adabiyotlar tahlili

So‘nggi yillarda Diabetes mellitus kasalligini erta aniqlash va prognoz qilish masalasi sun’iy intellekt va mashinali o‘rganish sohasida keng o‘rganilayotgan yo‘nalishlardan biriga aylandi. Tadqiqotchilar tomonidan turli algoritmlar yordamida diabetni aniqlash samaradorligini oshirishga qaratilgan ko‘plab ilmiy ishlar amalga oshirilgan.

Ko‘pgina tadqiqotlarda Pima Indians Diabetes Database asosiy ma’lumotlar manbasi sifatida qo‘llaniladi. Ushbu datasetning keng qo‘llanishi uning standartlashtirilganligi, ochiq mavjudligi va klinik parametrlarning yetarliligi bilan izohlanadi. Tadqiqotlar shuni ko‘rsatadiki, mazkur ma’lumotlar to‘plami diabetni bashorat qilish modellari uchun benchmark sifatida xizmat qiladi.

Ilk tadqiqotlarda asosan klassik statistik metodlar, xususan logistika regressiya modeli qo‘llanilgan. Ushbu yondashuv oddiyligi va interpretatsiya qilish qulayligi bilan ajralib turadi. Biroq, murakkab va nolinear bog‘liqliklarni aniqlashda uning imkoniyatlari cheklanganligi sababli, keyinchalik yanada rivojlangan algoritmlar qo‘llanila boshlandi.

Qaror daraxtlari va tasodifiy o‘rmon (Random Forest) algoritmlari diabetni aniqlashda keng qo‘llaniladigan metodlardan biri hisoblanadi. Tadqiqotlar natijalariga ko‘ra, Random Forest algoritmi yuqori aniqlik va barqarorlikni ta’minlaydi, chunki u bir nechta daraxtlar asosida qaror qabul qiladi va overfitting muammosini kamaytiradi. Shu bilan birga, ushbu model muhim atributlarni aniqlash imkonini ham beradi.

Support Vector Machine (SVM) algoritmi ham diabetni klassifikatsiya qilishda samarali usullardan biri sifatida qayd etilgan. Ushbu metod yuqori o‘lchamli ma’lumotlar bilan ishlashda yaxshi natija beradi va klasslar orasidagi optimal chegara (hyperplane)ni aniqlash orqali yuqori aniqlikni ta’minlaydi. Biroq, uning hisoblash murakkabligi katta ma’lumotlar to‘plamida muammo tug‘dirishi mumkin.

So‘nggi tadqiqotlarda chuqur o‘rganish (Deep Learning) metodlari, xususan sun’iy neyron tarmoqlar keng qo‘llanilmoqda. Ushbu yondashuv murakkab nolinear bog‘liqliklarni aniqlashda yuqori samaradorlikni namoyon etadi. Ayrim ilmiy ishlarda

neyron tarmoqlar yordamida 85–90% gacha aniqlik ko‘rsatkichlariga erishilgan. Biroq, bu metod katta hajmdagi ma’lumotlar va yuqori hisoblash resurslarini talab qiladi.

Bundan tashqari, ansambl metodlar, ya’ni bir nechta algoritmlarni birlashtirish orqali ishlovchi modellarning samaradorligi ham ko‘plab tadqiqotlarda isbotlangan. Masalan, Gradient Boosting va XGBoost algoritmlari diabetni aniqlashda yuqori natijalarni ko‘rsatgan. Ushbu metodlar xatoliklarni bosqichma-bosqich kamaytirish orqali model aniqligini oshiradi.

Adabiyotlar tahlili shuni ko‘rsatadiki, har bir algoritmning o‘ziga xos afzallik va kamchiliklari mavjud. Klassik metodlar oddiy va tez ishlaydi, biroq murakkab bog‘liqliklarni aniqlashda cheklangan. Chuqur o‘rganish metodlari esa yuqori aniqlik beradi, ammo katta resurs talab qiladi. Shu sababli, optimal modelni tanlash tadqiqot maqsadi va mavjud ma’lumotlarga bog‘liq.

Mazkur tadqiqot mavjud ilmiy ishlarni tahlil qilgan holda, diabet kasalligini aniqlash uchun turli mashinali o‘rganish algoritmlarini qo‘llash va ularning samaradorligini o‘zaro solishtirishga qaratilgan. Bu esa eng optimal va amaliy jihatdan samarali modelni aniqlash imkonini beradi.

Metodologiya

Mazkur tadqiqotda Diabetes mellitus kasalligini erta aniqlash uchun mashinali o‘rganish algoritmlariga asoslangan model ishlab chiqildi. Tadqiqot jarayoni bir nechta bosqichlardan iborat bo‘lib, ular ma’lumotlarni yig‘ish, oldindan qayta ishlash, model qurish, baholash va natijalarni tahlil qilishni o‘z ichiga oladi.

Ma’lumotlar to‘plami

Tadqiqot uchun Pima Indians Diabetes Database dan foydalanildi. Ushbu ma’lumotlar to‘plami tibbiy ko‘rsatkichlarga asoslangan bo‘lib, unda bemorlarning diabetga chalinish ehtimolini aniqlash uchun muhim bo‘lgan atributlar mavjud.

Dataset quyidagi asosiy parametrlarni o‘z ichiga oladi:

- Homiladorliklar soni
- Qondagi glyukoza darajasi
- Qon bosimi

- Teri qalinligi
- Insulin miqdori
- Tana massasi indeksi
- Genetik ko‘rsatkich (Diabetes Pedigree Function)
- Yosh

Natijaviy ustun esa bemorda diabet mavjud yoki yo‘qligini bildiradi.

Ma’lumotlarni oldindan qayta ishlash

Ma’lumotlar sifatini oshirish model aniqligiga bevosita ta’sir qiladi. Shu sababli quyidagi preprocessing bosqichlari amalga oshirildi:

- Nol qiymatlarni aniqlash va ularni o‘rtacha qiymatlar bilan to‘ldirish
- Ma’lumotlarni normalizatsiya qilish
- Atributlar orasidagi korrelyatsiyani tahlil qilish
- Trening va test to‘plamlariga ajratish (80% va 20%)

Model qurish

Tadqiqotda bir nechta mashinali o‘rganish algoritmlari qo‘llanildi va ularning samaradorligi solishtirildi:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine

Logistik regressiya modelida ehtimollik quyidagi formula orqali aniqlanadi:

$$P(y=1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}$$

Ushbu model bemorning diabetga chalinish ehtimolini hisoblashda ishlatiladi.

Random Forest algoritmi esa bir nechta qaror daraxtlari natijalarini birlashtirish orqali yakuniy qarorni qabul qiladi. Bu esa modelning aniqligini oshirish va overfitting muammosini kamaytirishga yordam beradi.

Support Vector Machine algoritmi esa ma’lumotlarni ikki sinfga ajratishda optimal chegara chizig‘ini aniqlaydi va yuqori aniqlikni ta’minlaydi.

Modelni baholash

Modellar quyidagi metrikalar yordamida baholandi:

- Accuracy
- Precision
- Recall
- F1-score

Baholash natijalari modelning samaradorligini aniqlash va eng optimal algoritmni tanlash imkonini berdi.

Tadqiqot jarayoni

Umumiy jarayon quyidagicha amalga oshirildi:

1. Ma’lumotlarni yuklash va tahlil qilish
2. Preprocessing bosqichlarini bajarish
3. Modellarni o‘qitish
4. Test ma’lumotlarda tekshirish
5. Natijalarni solishtirish va tahlil qilish

Mazkur metodologiya diabet kasalligini aniqlashda sun’iy intellektdan foydalanishning amaliy modelini yaratishga imkon berdi.

Natijalar va tahlil

Mazkur tadqiqotda Pima Indians Diabetes Database asosida qurilgan mashinali o‘rganish modellari sinovdan o‘tkazildi va ularning samaradorligi turli baholash metrikalari orqali aniqlandi. Tadqiqotning asosiy maqsadi Diabetes mellitus kasalligini yuqori aniqlik bilan bashorat qiluvchi optimal modelni aniqlashdan iborat edi.

Model natijalari

Tajriba natijalariga ko‘ra, har bir algoritm turli darajada samaradorlik ko‘rsatdi. Quyidagi jadvalda asosiy metrikalar bo‘yicha natijalar keltirilgan:

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Logistic Regression	0.76	0.74	0.72	0.73
Decision Tree	0.70	0.68	0.69	0.68
Random Forest	0.84	0.82	0.80	0.81
SVM	0.79	0.77	0.75	0.76

Natijalardan ko‘rinib turibdiki, Random Forest algoritmi eng yuqori aniqlikni ta’minladi. Bu uning bir nechta qaror daraxtlari asosida ishlashi va ma’lumotlardagi murakkab bog‘liqliklarni aniqlash qobiliyati bilan izohlanadi.

Muhim atributlar tahlili

Model natijalari asosida quyidagi parametrlar diabetni aniqlashda eng muhim omillar sifatida ajralib chiqdi:

- Glyukoza darajasi
- Tana massasi indeksi
- Yosh
- Insulin miqdori

Ayniqsa, glyukoza darajasi eng kuchli ta’sir qiluvchi omil sifatida qayd etildi. Bu tibbiy jihatdan ham asoslangan natija hisoblanadi.

Model solishtirish

Turli algoritmlarni solishtirish natijasida quyidagi xulosalarga kelindi:

- Logistic Regression oddiy va tez ishlaydi, lekin murakkab bog‘liqliklarni to‘liq aks ettira olmaydi
- Decision Tree interpretatsiya qilish uchun qulay, ammo aniqligi pastroq
- Random Forest eng barqaror va yuqori aniqlik beradi
- SVM muvozanatli natija beradi, lekin parametr sozlashga sezgir

Natijalar muhokamasi

Olingan natijalar shuni ko‘rsatadiki, sun’iy intellekt asosidagi modellardan foydalanish diabet kasalligini erta aniqlashda samarali yondashuv hisoblanadi. Ayniqsa, ansambl metodlar real tibbiy tizimlarda qo‘llash uchun eng istiqbolli yechimlardan biri sifatida namoyon bo‘ldi.

Biroq, modelning aniqligi ishlatilgan ma’lumotlar sifati va hajmiga bog‘liq. Mazkur tadqiqotda foydalanilgan Pima Indians Diabetes Database nisbatan kichik bo‘lgani sababli, natijalarni yanada yaxshilash uchun katta va real klinik ma’lumotlar bilan ishlash tavsiya etiladi.

Xulosa

Mazkur tadqiqotda Diabetes mellitus kasalligini erta aniqlash uchun sun’iy intellekt asosidagi modellarning samaradorligi tahlil qilindi. Tadqiqot davomida Pima

Indians Diabetes Database ma’lumotlar to‘plami asosida bir nechta mashinali o‘rganish algoritmlari ishlab chiqildi va ularning natijalari o‘zaro solishtirildi.

Olingan natijalar shuni ko‘rsatdiki, Random Forest algoritmi eng yuqori aniqlik ko‘rsatkichiga ega bo‘lib, diabetni bashorat qilishda eng samarali model sifatida ajralib chiqdi. Ushbu model ma’lumotlar orasidagi murakkab bog‘liqliklarni aniqlashda yuqori imkoniyatlarga ega ekanligi bilan ajralib turadi. Shu bilan birga, Logistic Regression va SVM algoritmlari ham barqaror natijalar ko‘rsatib, amaliy jihatdan foydali ekanligi aniqlandi.

Tadqiqot natijalari sun’iy intellekt texnologiyalaridan foydalanish orqali diabet kasalligini erta aniqlash imkoniyatlari mavjudligini tasdiqladi. Bu esa kasallikning og‘ir asoratlarini oldini olish, davolash jarayonini optimallashtirish va sog‘liqni saqlash tizimida samaradorlikni oshirishga xizmat qiladi.

Shu bilan birga, tadqiqotda ayrim cheklovlar mavjud bo‘lib, ular natijalarning umumlashuv darajasiga ta’sir qilishi mumkin. Xususan, ishlatilgan dataset hajmining cheklanganligi va real klinik sharoitlardan farqlanishi modelning amaliy qo‘llanilishida qo‘shimcha tadqiqotlar zarurligini ko‘rsatadi.

Ushbu ilmiy ish sun’iy intellekt asosida diabet kasalligini aniqlash bo‘yicha samarali yondashuvni taklif etadi. Tadqiqot natijalari kelajakda aqlli tibbiy diagnostika tizimlarini yaratish uchun asos bo‘lib xizmat qilishi mumkin. Shuningdek, mazkur yondashuv boshqa kasalliklarni aniqlashda ham qo‘llanishi mumkin bo‘lgan universal model sifatida qaraladi.

Foydalanilgan adabiyotlar

1. International Diabetes Federation. IDF Diabetes Atlas, 10th Edition
2. Diabetes mellitus. Available at https://en.wikipedia.org/wiki/Diabetes_mellitus
3. Pima Indians Diabetes Database. Kaggle platform. Available at https://www.kaggle.com
4. Scikit-learn documentation. Available at https://scikit-learn.org
5. TensorFlow documentation. Available at https://www.tensorflow.org
6. Patel, V., et al. Artificial intelligence in diabetes diagnosis and prediction. Computer Methods and Programs in Biomedicine