

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Сумський державний університет
Факультет електроніки та інформаційних технологій
Кафедра комп'ютерних наук

«До захисту допущено»

В.о. завідувача кафедри

_____ Ігор ШЕЛЕХОВ
(підпис)

09 червня 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА
на здобуття освітнього ступеня бакалавр

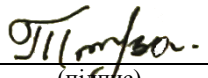
зі спеціальності 122 – Комп'ютерних наук,

освітньо- професійної програми «Інформатика»

на тему: «ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ЛОКАЛІЗУВАННЯ ПАТОЛОГІЇ В
ТКАНИНАХ МОЛОЧНОЇ ЗАЛОЗИ»

здобувача групи ІН-04р, Турапова Оразгул

Кваліфікаційна робота містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання на
відповідне джерело.

 Турапова Оразгул
(підпис)

Керівник
доцент,
кандидат технічних наук

Альона МОСКАЛЕНКО _____
(підпис)

Суми – 2024

Сумський державний університет
Факультет електроніки та інформаційних технологій
Кафедра комп'ютерних наук

«Затверджую»

В.о. завідувача кафедри

Ігор ШЕЛЕХОВ

(підпис)

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня бакалавра

зі спеціальності 122 - Комп'ютерних наук, освітньо-професійної програми «Інформатика»
здобувача групи Ін-04р, Турапова Оразгул

1. Тема роботи: «Інформаційна технологія локалізування патології в тканинах молочної залози»

затверджую наказом по СумДУ від «01» червня 2024 р. № 0475-VI

2. Термін здачі здобувачем кваліфікаційної роботи до 09 червня 2024 року

3. Вхідні дані до кваліфікаційної роботи _____

4. Зміст розрахунково-пояснювальної записки (перелік питань, що їх належить розробити)

1) Огляд медичних ознак при діагностуванні онкопатологій молочної залози 2) Огляд засобів інтелектуального аналізу онкопатологій тканин молочної залози. 3) Формування та аналіз вхідних даних. 4) Реалізація алгоритму сегментації тканини молочної залози

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) _____

6. Консультанти до проєкту (роботи), із значенням розділів проєкту, що стосується їх

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання « ____ » _____ 20 ____ р.

Завдання прийняв до виконання

Т. Шелехов
(підпис)

Керівник

(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Термін виконання	Примітка
1	<i>Огляд медичних ознак при діагностуванні онкопатологій молочної залози</i>		
2	<i>Огляд засобів інтелектуального аналізу онкопатологій тканин молочної залози</i>		
3	<i>Формування та аналіз вхідних даних</i>		
4	<i>Реалізація алгоритму сегментації тканини молочної залози</i>		

Здобувач вищої освіти

(підпис)

Керівник

(підпис)

АНОТАЦІЯ

Записка: 39 стр., 9 рис., 3 таблиці, 1 додаток, 24 використаних джерел.

Обґрунтування актуальності теми роботи – Методи сучасної цифрової мікроскопії надають гістологічні знімки надвисокої роздільної здатності, які візуально складно аналізувати із-за надмірної кількості інформаційних ознак. Тому актуальним є створення системи підтримки прийняття рішень, яка б була здатна автоматизувати процес сегментації зображення шляхом виділення зони інтересу.

Об'єкт дослідження – процес підтримки прийняття рішень.

Мета роботи – розробка алгоритмічного комплексу для сегментації гістологічних зображень тканин молочної залози.

Методи дослідження – моделі та методи обробки зображення та комп'ютерного зору.

Результати – система підтримки прийняття рішень лікаря-онколога для сегментації та початкового локалізування онкології в тканинах молочної залози

ГІСТОЛОГІЯ, ІНТЕЛЕКТУАЛЬНА СИСТЕМА, ІНФОРМАЦІЙНА
ПІДТРИМКА, МАШИННЕ НАВЧАННЯ, МОЛОЧНА ЗАЛОЗА,
КЛАСИФІКАЦІЯ, ОНКОПАТОЛОГІЯ

ЗМІСТ

Вступ	5
1	7
1.1	7
1.2	16
1.3	25
2	27
2.1	27
3	32
3.1	32
ВИСНОВКИ	33
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	34
ДОДАТОК	37

1. ВСТУП

Онкологічні захворювання, за статистикою, є другою із основних причин смертності у світі. Відомо, що виявлення онкозахворювання на ранніх стадіях суттєво підвищує успішність його лікування. Основним методом діагностування злоякісних пухлин є візуальний аналіз гістологічних зображень, отриманих за допомогою оптичного мікроскопу. При цьому достовірність діагнозу суттєво залежить від рівня професіональної підготовки та досвіду лікаря-гістолога. Так за даними Всесвітньої організації охорони здоров'я у 12 пацієнтів із 100 помилковий діагноз приймається або з помилкою першого роду, що обумовлює невиправдане лікування, або з помилкою другого роду - пропуск злоякісного новоутворення на ранній стадії його розвитку. Проте, сучасне обладнання дозволяє за методом повнослایдової цифрової мікроскопії отримувати високоякісні гістологічні зображення морфології тканини, розміщеної на предметному склі. З одного боку отримане за результатом сканування кольорове зображення великого розміру містить в собі кількість інформації, достатню для постановки правильного діагнозу. З іншого боку для лікаря достатньо важко візуально аналізувати повнослайдові зображення через багатовимірність словника діагностичних ознак. Тому, створення для лікаря-гістолога системи підтримки прийняття рішень, основною функцією якої є первинний аналіз вхідних даних, є актуальною задачею. При цьому зрозуміло, що юридичну відповідальність за достовірність несе безпосередньо лікар, який ставить остаточний діагноз.

Під час бакалаврського дослідження, буде розглянута задача інформаційного синтезу здатної навчатися системи діагностування онкопатологій молочної залози за гістологічними зображеннями

1 ІНФОРМАЦІЙНИЙ ОГЛЯД

1.1 Інформаційний огляд медичних ознак при діагностуванні онкопатологій молочної залози.

Згідно інформації Всесвітньої організації охорони здоров'я рак молочної залози є найбільш розповсюдженим онкозахворюванням серед громадян України [1]. При цьому, відповідно до того ж самого джерела, цей недуг, станом на 2021 рік, є найбільш розповсюдженим типом онкопатологій в світі [2]. Незважаючи на розвиток методів діагностування, людські ресурси все ще залишаються обмеженими. У більшості країн лікарі онкологи фізично не здатні забезпечити лікуванням достатню кількість хворих. Тому створення системи підтримки прийняття рішень для діагностування онкопатологій є актуальною задачею. Основною складністю, з якою зіштовхується будь-який лікар в медичній галузі, є індивідуальні біомаркери кожного окремого організму. Вони впливають на внутрішні процеси та, як наслідок, можуть формувати не типові симптоми під час захворювання та в процесі лікування. Так, наприклад, кожного дня в онко-лабораторіях по всьому світу знаходять нові типи новоутворень, які можуть відрізнятися за формою, щільністю чи темпами розповсюдження. Саме тому більшість лікарів вважають, що кожен організм потрібно комплексно розглядати і відокремлювати його від інших. Цей напрям досліджень називається персоналізована медицина.

При діагностуванні онкопатологій молочної залози використовуються два принципово різних підходи: рентгенівський скринінг (мамографія) і оптична мікроскопія. Кожен із цих методів має свої переваги та недоліки, які потрібно враховувати при проектуванні інтелектуальної системи.

Цифрова маммографія молочної залози – є неінвазивним методом діагностичного обстеження молочної залози, що дозволяє виявити злоякісні новоутворення на різних стадіях патологічного процесу. Застосовується як основний метод або як додатковий для інших способів діагностики: рентгену, магнітно-резонансної томографії (МРТ), комп'ютерної томографії (КТ).

Приклад діагностування раку молочної залози за її рентгенівським знімком наведений в праці [1].

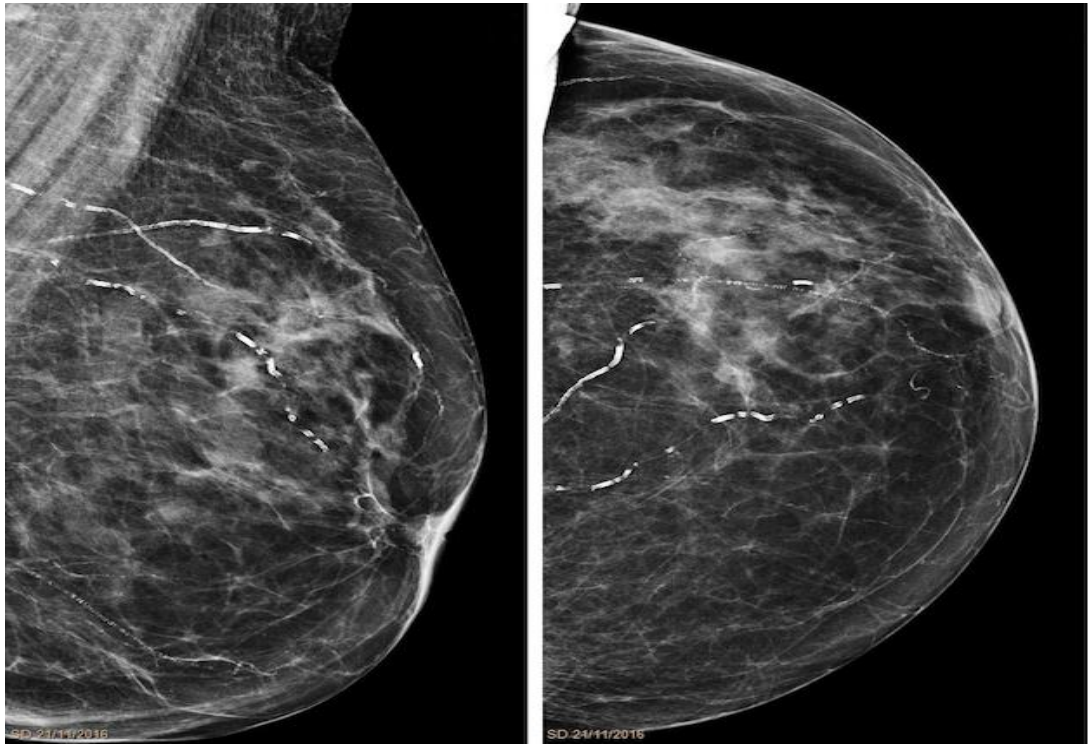


Рисунок 1.1 – Приклад зображення цифрової мамографії

Хоча дослідження патологій тканин шляхом магнітно-резонансної томографії є високо достовірним та дуже поширеною процедурою діагностування онкологій найрізноманітнішого генезу. Більше того, знімки МРТ є досконально дослідженими та строго формалізованими. Це, перш за все, й обумовлює популярність цього методу діагностування. Тим не менш, кількість інформації, яку лікар може отримати з цього знімку (рис.1.1.) також строго обмежена. Є можливість локалізувати новоутворення, оцінити темпи його розвитку, але не можливо врахувати цитологічні біомаркери пацієнта або ж виявити гістологічні ознаки, які характерні для різних форм раку. Це обумовлено самою специфікою знімку, рисунок 1.1., який не може містити в собі подібної інформації.

Таким чином складність діагностування тканин за їх рентгенівськими знімками ускладнюється наступними факторами: По-перше, ракові пухлини є не стаціонарними за своєю формою. По-друге, на розвиток онкопатології

впливають різноманітні біомаркери, які є індивідуальними для кожної людини. Досить цікавий факт, що майже кожного дня в медичних лабораторіях по всьому світу знаходять нові види злоякісних новоутворень, які відрізняються своєю формою, щільністю чи іншими характеристиками. Приклад діагностування раку молочної залози за її рентгенівським знімком наведений в праці [3]. Головним недоліком такого виду діагностування, з точки зору проектування інтелектуальної системи, є брак інформативних ознак.

Другий метод, який використовують лікарі для діагностування онкозахворювань, є оптична мікроскопія. Цей підхід надає набагато більше інформації про вражену тканину та дозволяє набагато точніше класифікувати тип злоякісного новоутворення. У сучасній медицині набуває популярності метод, так званої, повнослایдової цифрової мікроскопії (whole-slide-imaging, WSI) [3, 4]. У рамках цієї технології виконується зйомка усієї тканини, розміщеної на предметному склі (рис. 1.2).

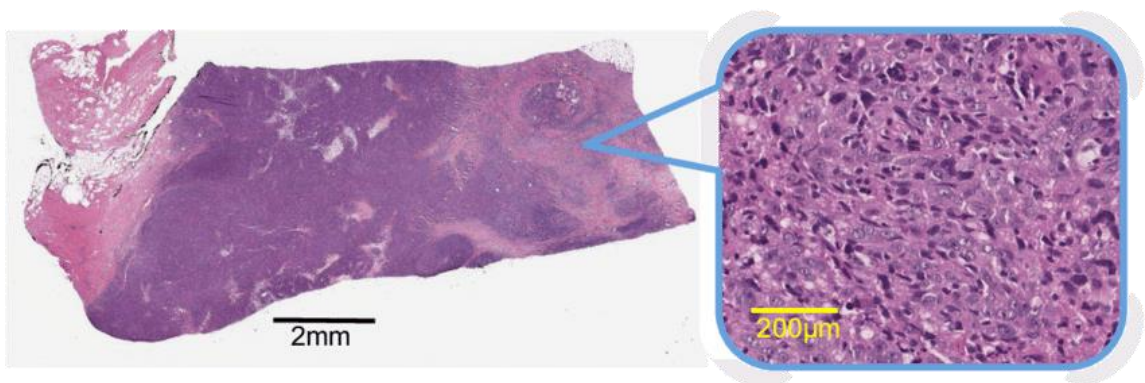


Рисунок 1.2 – Приклад повнослایдового зображення

Незважаючи на те, що метод WSI дозволяє отримати, достатню кількість інформації для постановки правильного діагнозу, проте, без належного програмного забезпечення, використання повнослایдових зображень є ускладненим. Це справедливо помітили доктор В.Луо та професор Л.Хассел у своїй праці [3]. Проблема полягає у тому, що гістологічні знімки мають надмірну роздільну здатність. Це спричиняє складнощі для лікаря при візуальному аналізі

тканини. Річ у тім, що більшість форм патологій мають мультицентричну природу. Це означає, що розвиток раку починається одразу в декількох місцях. Із-за цього знайти усі його осередки на ранній стадії розвитку може бути достатньо складно. Також варто мати на увазі, що далеко не кожна патологія є злоякісною, тому при аналізі гістологічних знімків (рис.1.2.) лікарю необхідно спочатку локалізувати ураження, а вже потім його діагностувати, що подекуди неможливо зробити без урахування суміжних, з осередком захворювання, тканин.

Класифікацію зображення тканини можна розглядати як процес відділення ділянок, які вражені злоякісними новоутвореннями, від здорових. Більшість інтелектуальних систем діагностування, які описано в сучасній літературі, можна розділити на три класи:

1) Системи, які базуються на графічних методах обробки зображення. Як один із варіантів, сегментацію новоутворень можна реалізувати через послідовне використання алгоритму покадрової заливки зображення. При цьому для кожного знайденого об'єкту інтересу буде застосовано метод дилатації, який визначить його межі. Фактично, ці системи не є інтелектуальними, бо в них відсутній етап машинного навчання та прийняття рішень. Очевидним недоліком цього підходу є залежність від оператора, який виконує перетворення зображення та оцінює отриманий результат. Окрім цього ці системи характеризуються високою вірогідністю прийняття помилки першого роду, що обумовлює невиправдане лікування, або помилки другого роду – пропуск злоякісного новоутворення на ранній стадії його розвитку.

2) Системи, що базуються на методах «не глибокого» навчання. Зазвичай вхідними даними алгоритмів цього класу є геометричні параметри попередньо сегментованих об'єктів. До цієї групи можна віднести: алгоритм випадкових лісів, метод опорних векторів, найпростіші штучні нейронні мережі (ШНМ) тощо.

3) Системи, які здійснюють глибоке навчання. Їх головною особливістю є комплексний аналіз зображення. До цього класу можна віднести згорткові ШНМ.

Перспективним напрямом є створення СППР для лікаря-онколога. Для вирішення цього питання, у медичній галузі, широкого поширення набули ШНМ [5, 6]. Нажаль цей підхід має ряд недоліків, які проявляються під час обробленні великих масивів даних, які сформовані при довільних умовах.

Перед початком розробки інтелектуальної компоненти діагностування раку молочної залози необхідно проаналізувати не лише вхідні дані, у вигляді знімків тканини, а й усі попередні діагностичні дії, які виконує лікар. Адже чим краще буде досліджений процес прийняття рішень, тим більш досконалий програмний застосунок вдасться реалізувати.

Після того, як у когось діагностовано рак молочної залози, лікарі спробують з'ясувати, чи поширився він, і якщо так, то як далеко. Цей процес називається постановкою стадії раку, який визначає, скільки вражених клітин в організмі. Це допомагає визначити, наскільки серйозним є патологія і як краще її лікувати. Лікарі також корелюють дані класифікації раку з статистикою летальності.

Найпершою стадією раку молочної залози є 0 (карцинома). У подальшому він варіюється від стадії I(1) до IV(4). Як правило, що менше число, то менше поширився рак. Вища кількість, наприклад, стадія IV, означає, що патологія захопила більшу кількість тканин. Хоча досвід лікування раку у кожної людини унікальний, розвиток патології, зазвичай, відповідає її класифікації. Тому можна стверджувати, що люди з, наприклад, III стадією раку будуть мати схожі симптоми, а отже й подібне лікування.

Системою стандартизації, що найчастіше використовується при раку молочної залози, є TNM Американського об'єднаного комітету з раку (AJCC). Остання система AJCC, що діє з січня 2018 року, включає як клінічну, так і патологічну систему класифікації патологій тканин молочної залози:

- **Патологічна стадія** (також звана хірургічною) визначається шляхом дослідження зразків тканини видалених під час операції.
- Іноді, якщо операція неможлива відразу або взагалі, тоді раку присвоюється **клінічна стадія**, яка ґрунтується на результатах фізичного огляду, біопсії та візуалізуючих тестів. Клінічний етап використовується для планування лікування. Однак іноді рак поширився далі, ніж передбачалося на клінічній стадії. Для пацієнта це означає, що неможливо дати достовірний прогноз стосовно перебігу хвороби.

Під час дослідження паталогії молочної залози лікарі досліджують наступні ознаки тканини:

- Ступінь (розмір) пухлини (T): Наскільки великі ракові залози? Чи відбувається інвазія в сусідні тканини?
- Поширення на прилеглі лімфатичні вузли (N): Чи поширився патологічний процес на прилеглі лімфатичні вузли? Якщо так, то наскільки?
- Поширення (метастазування) на віддалені ділянки (M): Чи поширився рак на віддалені органи, такі як легені чи печінка?
- Статус рецептора естрогену (ER): чи виробляють ракові залози білок, який має назву рецептором естрогену?
- Статус рецептора прогестерону (PR): чи виробляють ракові залози білок, який називається рецептором прогестерону?
- Статус HER2: чи виробляють ракові залози занадто багато білка під назвою HER2?
- Ступінь раку (G): Наскільки ракові клітини схожі на нормальні клітини?

Іноді до цього списку додають оцінку можливості рецидиву, коли пацієнт вже проходив лікування від патологій молочної залози або інших органів. Також можуть враховуватися генетичні особливості, наприклад, коли родичі пацієнта

також мали онкологічні захворювання. Так чи інакше, ці два фактори є ситуативними.

Після визначення усіх доступних ознак ця інформація об'єднується у процесі, що називається групуванням стадій, для призначення загального етапу патології. Не зважаючи на те, що при прийнятті рішень лікар використовує велику кількість інформації, проте специфіка тканин молочної залози дозволяє класифікувати стадію патології за допомогою візуального аналізу пухлини (tumor, T), стану лімфатичних вузлів (lymph node, N) та метастазами (metastasis, M). У медичній літературі цей метод діагностування має назву система TNM.

Категорія T для раку молочної залози оцінюється від 0 до 4 й описує розмір основної (первинної) пухлини, а також враховує, чи вона поширилася на шкіру або на стінку грудної клітки під грудьми.

Таблиця 1.1. Категорії стану пухлини T

Назва категорії	Опис
TX	Первинну пухлину оцінити неможливо
T0	Немає ознак первинної пухлини
Tis	Карцинома in situ (DCIS або хвороба Педжету молочної залози без супутньої пухлинної маси).
T1 (T1a, T1b, T1c)	пухлина має розмір 2 см (3/4 дюйма) або менше в поперечнику.
T2	пухлина більше 2 см, але не більше 5 см (2 дюйми) у поперечнику.
T3	Пухлина більше 5 см у поперечнику.
T4 (T4a, T4b, T4c та T4d)	пухлина будь-якого розміру, що проростає в грудну стінку або шкіру. Сюди входить запальний рак молочної залози.

Категорія N визначається від 0 до 3 та вказує, чи поширився рак на лімфатичні вузли поруч із грудьми, і якщо так, то наскільки вони вражені. Оцінюється те, як вони виглядають на скринінгу. При цьому, методи сучасної цифрової мікроскопії дозволяють максимально досконально проаналізувати ці тканини, майже на рівні клітин. Тим не менш, багато лікарів не впевнені чи виправдана така деталізація при діагностування лімфатичних вузлів.

Сучасним науковцям все ще не зрозуміло, який рівень ураження лімфатичного вузла необхідний, щоб побачити зміни в прогнозі або лікуванні. Це питання все ще досліджується, але на даний момент депозит ракових клітин повинен містити не менше 200 або мати діаметр не менше 0,2 мм (менше 1/100 дюйма), щоб він міг змінити стадію N. Область поширення раку розміром менше 0,2 мм (або менше 200 клітин) не змінює стадію, а позначається скороченнями (i+ або моль+), що вказує на тип спеціального тесту, використовуваного при виявленні поширення.

Якщо площа поширення раку становить не менше 0,2 мм (або 200 клітин), але все ж таки не перевищує 2 мм, це називається мікрометастазом (один мм приблизно дорівнює ширині рисового зернятка). Мікрометастази враховуються тільки в тому випадку, якщо немає більших осередків розповсюдження раку. Відомо, що ділянки поширення раку розміром більше 2 мм впливають на зовнішній вигляд і змінюють стадію N. Ці більші області іноді називають макрометастазами, але найчастіше їх називають просто метастазами.

Таблиця 1.2. Категорії стану лімфовузлів N

Назва категорії	Опис
NX	Неможливо оцінити довколишні лімфатичні вузли (наприклад, якщо вони були видалені раніше).
N0	Рак не поширився на довколишні лімфатичні вузли.
N0(i+)	Область поширення раку містить менше 200 клітин та має розмір менше 0,2 мм. Аббревіатура «i+» означає, що невелика кількість ракових клітин (званих ізольованими пухлинними клітинами) була виявлена при звичайному фарбуванні або при використанні спеціального методу фарбування, що називається імуногістохімією.
N0(моль+)	Ракові клітини неможливо виявити в пахвових лімфатичних вузлах (навіть при використанні спеціальних барвників), але сліди ракових клітин були виявлені за допомогою методу, званого ВІД-ПЛР. RT-PCR – це молекулярний тест, який дозволяє виявити дуже невелику кількість ракових клітин.
N1	Рак поширився на 1–3 пахвових лімфатичних вузла(ів) та/або рак виявляється у внутрішніх лімфатичних вузлах молочної залози (біля грудини) при біопсії сторожового лімфатичного вузла.

N1mi	мікрометастази (невеликі ділянки поширення раку) у лімфатичних вузлах під пахвою. Ділянки поширення раку в лімфатичних вузлах становлять не менше 0,2 мм у поперечнику, але не більше ніж 2 мм.
N1a	Рак поширився на 1–3 лімфатичні вузли під пахвою, при цьому принаймні одна ділянка раку поширилася більше 2 мм у поперечнику.
N1b	Рак поширився на внутрішні лімфатичні вузли молочної залози на тій же стороні, що й рак, але це поширення можна

Таблиця 1.2. Категорії стану лімфовузлів N

	було виявити тільки за біопсії сторожового лімфатичного вузла (це не призвело до збільшення лімфатичних вузлів).
N2	рак поширився на 4-9 лімфатичних вузлів під пахвою або рак збільшив внутрішні лімфатичні вузли молочної залози.
N2a	рак поширився на 4-9 лімфатичних вузлів під пахвою, при цьому принаймні одна ділянка раку поширилася більш ніж на 2 мм.
N2b	Рак поширився на один або кілька внутрішніх лімфатичних вузлів молочної залози, що призвело до їх збільшення.
N3a	Рак поширився на 10 або більше пахвових лімфатичних вузлів, при цьому принаймні одна область раку поширилася більш ніж на 2 мм. АБО Рак поширився на лімфатичні вузли під ключицею (підключичні вузли), причому принаймні одна область раку поширилася більш ніж на 2 мм.
N3b	Рак виявляється принаймні в одному пахвовому лімфатичному вузлі (принаймні з однією областю поширення раку більше 2 мм) і призводить до збільшення внутрішніх лімфатичних вузлів молочної залози. АБО Рак поширився на 4 або більше пахвових лімфатичних вузлів (принаймні одна область раку поширилася більше 2 мм) і на внутрішні лімфатичні вузли молочної залози при біопсії сторожових лімфатичних вузлів.
N3c	Рак поширився на лімфатичні вузли над ключицею (надключичні вузли) на тій стороні від раку, при цьому принаймні одна область раку поширилася більш ніж на 2 мм.

Категорія M визначається як 0 або як 1, вказує, чи поширився рак на віддалені органи, наприклад легені, печінку або кістки.

Таблиця 1.3. Категорії стану поширення раку M

Назва категорії	Опис
-----------------	------

M0	При рентгенологічному дослідженні (або інших візуальних дослідженнях) або при фізичному огляді віддалене поширення не виявлено.
cM0(i+)	невелика кількість ракових клітин виявляється в крові або кістковому мозку (виявляється тільки за допомогою спеціальних тестів), або крихітні ділянки поширення раку (не більше 0,2 мм) виявляються в лімфатичних вузлах далеко від пахв, ключиць, або внутрішні області молочної залози.

Таблиця 1.3. Категорії стану поширення раку M

M1	Рак поширився на віддалені органи (найчастіше на кістки, легені, мозок або печінку), що видно при візуальних дослідженнях або фізичному огляді, та/або біопсія однієї з цих областей доводить, що рак поширився та його розміри перевищують 0,2 мм
----	--

Отже, при встановленні діагнозу пацієнта лікар використовує комплексну систему ознак, які характеризують рак. При цьому можна формалізувати патологічний процес наступними етапами:

- 1) Не примітна мутація залоз/клітин/тканин, яке не завжди переросте в злоякісне новоутворення.
- 2) Розвиток патологічного процесу у вигляді мультцентричного поширення, коли в суміжних тканинах з'являються схожі новоутворення.
- 3) Розростання та злиття уражених залоз.
- 4) Руйнування найбільших залоз та активне розповсюдження патології у суміжні тканини.

1.2 Інформаційний огляд засобів інтелектуального аналізу онкопатології тканин молочної залози

Функціональна ефективність сучасних інтелектуальних комп'ютеризованих систем все ще поступається природньому інтелекту людини. Тому задача моделювання когнітивних процесів, притаманних людині при формуванні та прийнятті класифікаційних рішень, є однією із найскладніших задач сучасності. Вирішення проблеми підвищення функціональної ефективності комп'ютеризованих систем діагностування, у тому числі, для

розпізнавання гістологічних зображень, можливо тільки шляхом застосування ідей і методів машинного навчання та розпізнавання образів. У загальному випадку процес розпізнавання образу складається з двох етапів:

1) Етап машинного навчання, на якому формуються деяким оптимальним способом вирішальні правила.

2) Етап безпосереднього розпізнавання, на якому за побудованими на етапі машинного навчання вирішальними правилами визначається належність реалізації образу, що розпізнається, одному із класів заданого алфавіту.

Основними підходами до вирішення проблем машинного навчання та розпізнавання образів є [8]:

1. Детермінований. Основна перевага: прості вирішальні правила. Основний недолік: низька достовірність розпізнавання, оскільки не враховуються додаткові неконтрольовані фактори, які впливають на процес розпізнавання.
2. Статистичний, в рамках якого використовуються статистичні вирішальні правила (наприклад, Байєсівський класифікатор та його модифікації). Основний недолік - необхідність для перевірки статистичної стійкості і статистичної однорідності вхідних даних використовувати репрезентативні навчальні вибірки великого обсягу.
3. Детерміновано-статистичний, який полягає в побудові детермінованих вирішальних правил, статистична корекція яких відбувається в процесі машинного навчання.
4. Геометричний, у рамках якого задача класифікації вирішується в N -вимірному просторі ознак, де N – фіксована потужність словника ознак розпізнавання. Якщо потужність алфавіту класів розпізнавання дорівнює M , то задача класифікації полягає в розбитті простору ознак на M класів розпізнавання. Саме в рамках геометричного підходу клас розпізнавання розглядається як область, в якій ознаки розпізнавання характеризуються відношенням еквівалентності.

5. Структурний (лінгвістичний), який моделює структуру об'єкту розпізнавання. У рамках цього підходу розробляються нейроподібні структури, семантичні мережі, фрейми, мережі Петрі, дерева рішень та інше. Переваги: простота моделі, можливість її розширення та ускладнення. Основними недоліками є:
 - складні вирішальні правила;
 - висока чутливість до багатовимірності простору ознак розпізнавання;
 - негнучкість, яка ускладнює перенавчання системи при розширенні алфавіту класів розпізнавання.
6. Функціональний, який моделює когнітивні процеси, притаманні людині при формуванні та прийнятті класифікаційних рішень.
7. Нечіткий. Цей підхід дозволяє моделювати важко формалізовані процеси, де ознаки розпізнавання визначаються за якісною шкалою виміру. Але при наявності кількісної шкали виміру ознак розпізнавання перевагу слід віддавати методам багатовимірного статистичного аналізу, які використовують більш розвинений математичний апарат.

Незважаючи на те, що наведені підходи відрізняються один від одного рівнем і видом математичної формалізації слабо формалізованих процесів прийняття рішень, між ними не існує чіткої межі, а самі підходи часто доповнюють один одного. Оскільки всі основні підходи, окрім алгебраїчного, перетинаються з геометричним, то саме в рамках геометричного підходу формування загальної теорії прийняття рішень є найбільш виправданим. Особливо перспективним є реалізація машинного навчання в рамках геометричного підходу при аналізі великих обсягів даних, що має місце і при використанні метода повнослайдової цифрової мікроскопії. Так, вирішальні правила, побудовані за отриманими в процесі машинного навчання геометричними параметрами, є практично інваріантні до багатовимірності, оскільки сучасні ЕОМ здатні обробляти структуровані вектори, які складаються

з неймовірно великого числа 285 ознак розпізнавання. Серед найбільш відомих комп'ютеризованих систем діагностування онкопатологій можна виділити систему діагностування раку молочної залози [9]. Вхідними даними для навчання і тестування були результати цитологічного дослідження тканини молочної залози, отриманої за методом біопсії. Структурований вектор складався всього з дев'яти характеристик клітин тканини молочної залози (товщина скупчення, однорідність розмір клітин, однорідність форми клітин, крайова адгезія, розмір одиничних епітеліальних клітин, ядро, блідість хроматину, нормальність ядерець, мітози) і одну змінну, яка фіксує доброякісність чи злоякісність пухлини. Набір включав 683 результати біопсії, з яких 444 зразки – тканини з доброякісними пухлинами, 239 – тканини зі злоякісними пухлинами. Для навчання використовували 200 зразків з доброякісними пухлинами та 200 зразків зі злоякісними пухлинами. Залишок – 283 зразки, які використовуються для тестування алгоритму. Оскільки база даних знаходилася у вільному доступі, то дослідники мали можливість створювати класифікаційні алгоритми різними методами. Найбільшу достовірність для алфавіту із двох класів було отримано при реалізації методу опорних векторів [10, 11]. Але варто зазначити, що при збільшенні потужності алфавіту класів розпізнавання, які перетинаються в просторі діагностичних ознак, збіжність методу опорних векторів не гарантується. Значного прогресу в області розпізнавання медичних зображень було досягнуто порівняно недавно у зв'язку з винайденням швидких графічних процесорів, які дають можливість будувати та навчати нейронні мережі з великою кількістю шарів [12]. З використанням нейроподібної структури відомий проект, реалізований компанією Google. Вхідними даними для навчання та тестування нейронної мережі був набір повнослайдових зображень Camelon 16 dataset, який складався з 400 мікроскопічних знімків тканин лімфовузлів. Як результат, найкращий варіант навченої нейронної мережі правильно ідентифікував 92,5% пухлин на тестовому наборі слайдів, а лікар-патологоанатом тільки 73,3%.

Проблему розпізнавання гістологічних зображень вирішують також українські науковці. Так, у праці [14] розглядалася задача розпізнавання трьох зображень морфології тканини, отриманої за методом біопсії для діагностування раку молочної залози. Автори отримали достатньо високу усереднену повну ймовірність правильного діагностування. Машинне навчання системи діагностування здійснювалося в рамках так званої інформаційно-екстремальної інтелектуальної технології (ІЕІ-технології) аналізу даних, яка ґрунтується на максимізації інформаційної спроможності системи в процесі її машинного навчання [16].

У праці [8] представлений приклад інтелектуальної компоненти, яка здатна розпізнавати ураження молочної залози за скринінгом цієї тканини. Вважається, що з інженерної точки зору це дещо простіша задача ніж гістологічний аналіз. Оскільки рентгенівські знімки містять обмежену кількість інформації про об'єкт інтересу. Тим не менш, автори стверджують, що точність їх системи підтримки прийняття рішень наближена не гірша, ніж у радіолога. Їх результати перевірялися на двох датасетах, які представляють різний скринінг населення та практики.

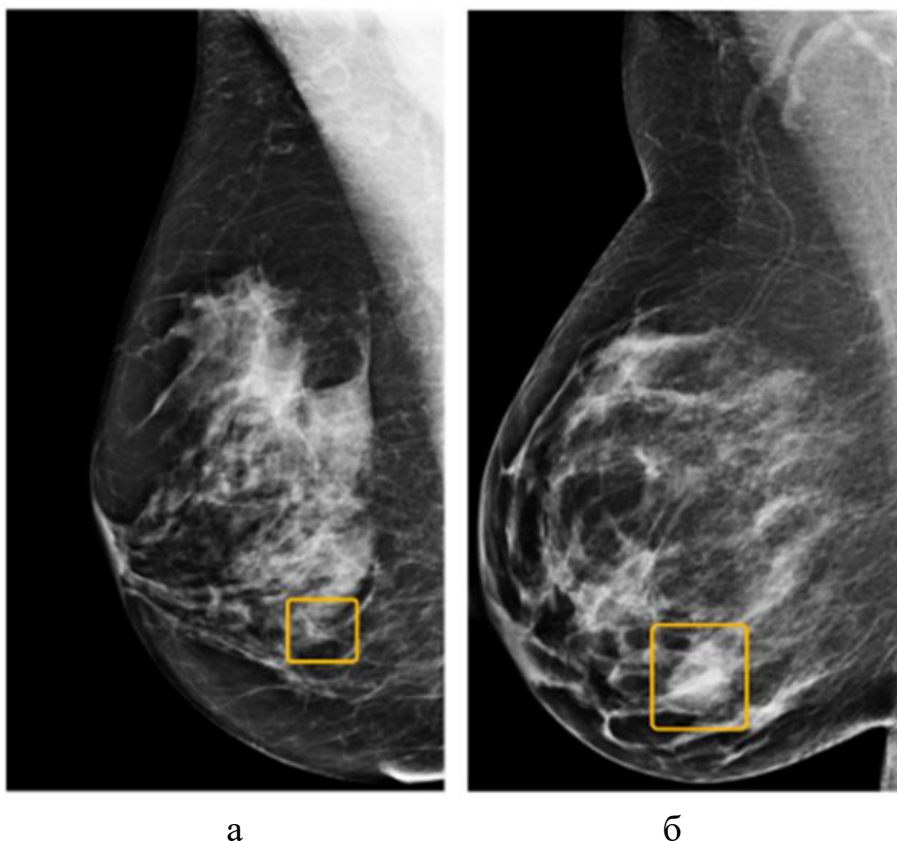


Рисунок 1.3. а) приклад мімікрії онкології під здорову тканину на ранній стадії розвитку раку; б) злоякісне новоутворення в тканинах молочної залози

На рисунку 1.3.б наведено ураження паталогією, яке пропустили всі шість лікарів-експертів у дослідженні авторів, але яке було правильно локалізоване системою ШІ. Злоякісна пухлина, обведена жовтим контуром, є невелике неправильне утворення з відповідними мікрокальцифікатами в нижній внутрішній частині тканини. При цьому на рисунку 1.3.а зразок випадку раку, який виявили всі шість експертів, але пропущене системою ШІ. Злоякісне новоутворення являє собою щільне скупчення залоз.

Варто зауважити, що у дослідженні [8] переважна більшість зображень, були отримані на пристроях виробництва Hologic. Цілком можливо, що розроблена інтелектуальна компонента здатна показувати високу ефективність лише при контрольованому вхідному математичному описі. Майбутні дослідження мають бути більш систематичними й оцінити продуктивність СППР для різних виробників пристрої МРТ.

Також важливо відмітити, що частково спеціалісти-радіологи, що були задіяні в дослідженні, навчалися та працюють в США, хоча самі знімки тканин були зроблені в Об'єднаному королівстві. Цілком можливо, що способи візуального аналізу та інтерпретації радіологічних знімків відрізняються в цих двох країнах.

У праці [9] автори, як вхідний математичний опис, використали гістологічні зображення тканини молочної залози. Ці знімки містять набагато більше інформативних ознак, порівняно з рентгенівськими. З одного боку це надає додаткових можливостей при дослідженні онкопатологій, а з іншої ускладнює його. Адже все ще не існує чіткого формалізованого списку гістологічних ознак, які б з сто відсотковою достовірністю вказували б на приналежність раку тій чи іншій стадії.

Окрім цього варто зауважити, що аналіз гістологічних зображень проводиться за результатами біопсії, яка є інвазивною медичною операцією. Це означає, що на цьому етапі лікарю вже відомо про наявність та локалізацію новоутворення в тканинах пацієнта. У більшості випадках, біопсія робиться не для підтвердження наявності патології в тканинах, а для прогнозування захворювання.

У своїй роботі [9] автори використовують повнослайдові гістологічні зображення. Оскільки вони мають надмірну роздільну здатність, то ціле зображення розбивається на невеликі частини, які стають вхідним описом для згорткової нейронної мережі (CNN). Невеликі частини початкового зображення назвемо патчами. Таким чином, CNN на рівні патчів, які класифікуються незалежно один від одного, генеруватимуть багато неправильних ізольованих ймовірностей.

Автори помітили закономірність поширення раку в тканинах, а саме його ізольовану природу. Візуально це виглядає як чітко виокремленні осередки патології, які оточені здоровими тканинами. Це означає, що патчи, які ідентифіковані як злякисне новоутворення повинні створювати конгломерат. З

точки зору ознак зображення, неправильні прогнози від CNN - це шум у карті ймовірностей, який необхідно видалити. З іншого боку, ядро фільтра повинно використовувати переваги розподілу міток сусідніх патчів для переоцінки виходів CNN. Як наслідок, модель не розглядає лише інформацію про текстуру, а й використовує мітки сусідніх патчів, отримуючи більш плавний розподіл ймовірностей.

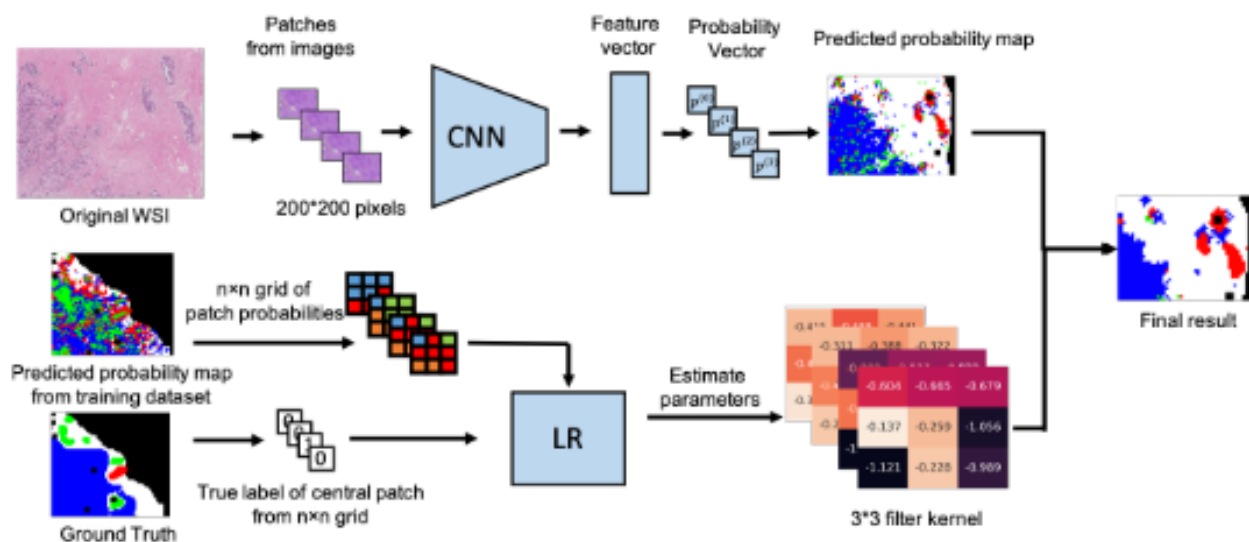


Рисунок 1.4. Схема основи для використання просторової інформації

Таким чином (рис. 1.4) автори пропонують двоетапну основу для класифікації патчів WSI. По-перше, вони навчають CNN для класифікації і використання мережі для вилучення ймовірності появи онкології в патчі. Це означає, що першому етапі загорткова мережа виступає в ролі як функція-екстрактор. Потім вони відфільтровують карту ймовірностей за допомогою ядра фільтра, щоб отримати кращий аналіз використовуючи переваги просторових кореляцій.

У загальному випадку робота [9] зосереджена на використанні просторових кореляцій між патчами. Адже хоч модель CNN здатна сама по собі виступати як класифікатор, але результат міститиме багато неправильних ізольованих ймовірнісних міток, оскільки ігноруються просторові кореляції.

Подальші експерименти показують, що збільшення розміру ядра може надати кращої продуктивності в класифікації, але точність стає меншою.

Подібну реалізацію запропонували й автори у роботі [11]. При роботі з WSI вони зосередились на локалізації ураження, а не на прогнозуванні хвороби. Так само, як і в [9] вони бінаризували зображення на здорову та уражену тканину. При цьому акцентується увага на метастазах, які контрастують з сполучною тканиною на гістологічних знімках.

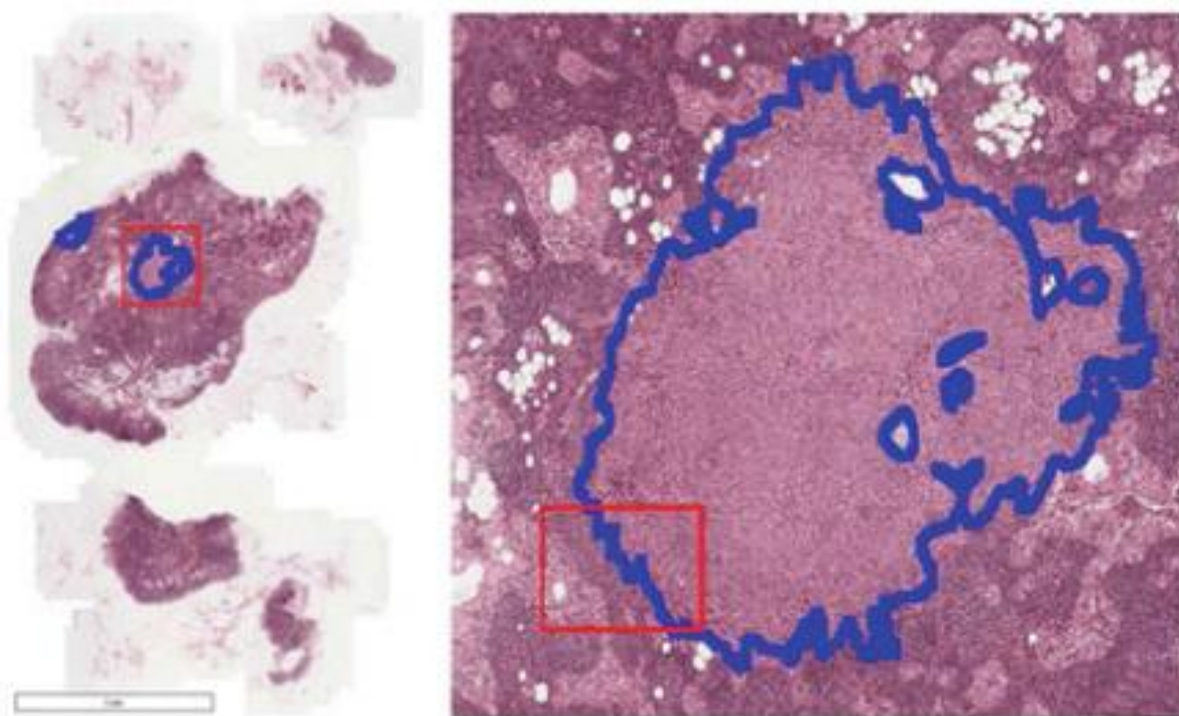


Рисунок 1.5. Результат локалізації ураження в тканинах молочної залози

Замість покращення вдосконалення алгоритму машинного навчання, у роботі [11] автори звертають свою увагу на попередній обробці вхідного математичного опису, для збільшення контрастності між здорової та враженими тканинами, зокрема, за допомогою методу селекції Otsu [12]. Це дозволяє зменшити кількість хибних ідентифікацій, що позитивно впливає на кінцевий результат.

Отже, основна технічна проблема при роботі з WSI зображеннями – це їх гігапиксельна природа. У більшості випадках не можливо завантажити увесь знімок одразу в пам'ять комп'ютера, щоб навчити систему розпізнавати його.

Для цього вхідний матеріал розбивається на патчі. У деякому сенсі, відбувається дослідження генеральної сукупності через множину її вибірок.

Окрім цього не існує єдиного стандарту діагностування гістологічних зображень. На практиці, кожен окремих лікар онколог, при прийнятті рішень, перш за все опирається на власну медичну практику. Таким чином, кінцевий діагноз, а отже і план лікування буде відрізнятись в залежності від лікаря. Це не означає, що фахівець не є компетентним в своїй галузі, а лише вказує на те, що його дії є результатом особистого досвіду, що цілком нормально. Проте, із-за цього можуть виникати розбіжності при діагностуванні, наприклад, для лікування доброякісного новоутворення. Там де один лікар буде настоювати на операції із-за потенційної загрози розвитку патології, інший вважатиме це мало ймовірним. Відсутність формалізованого списку онкопатологічних ознак з їх детальним описом та розвитком, ускладнює розробку системи підтримки рішень.

Окрім цього, достатньо не очевидною перешкодою при розробці СППР в медицині – є не вирішені етичні та юридичні питання [13]. Перш за все, будь-яка комп'ютеризована система не здатна нести відповідальність за помилкові рішення під час діагностування, які, так чи інакше, будуть траплятися на практиці. Обумовлено це тим, що на поточний момент часу не можливо створити максимально універсальний штучний інтелект. Так чи інакше, будь-який подібний алгоритм має слабкі місця. Зокрема вони чутливі до вхідних даних, які на практиці є довільними. Ці фактори ускладнюють повноцінну імплементацію інтелектуальних систем в роботу лікаря, де ціна помилки – це людське здоров'я та життя.

1.3 Постановка задачі

Необхідно розробити програмний комплекс для ідентифікації онкологічного процесу в тканинах молочної залози. При цьому варто звернути окрему увагу на подібність ранньої стадії раку на доброякісне новоутворення, що

обумовлено тісним перетином їх ознак-розпізнавання. А також на довільність умов формування повнослайдових знімків, які можуть відрізнятися в залежності від лабораторії, де ті були зроблені.

Як вхідний математичний опис пропонується використати знімки тканини молочної залози з проявом ранньої стадії раку у вигляді гіперхромії окремих залоз. Такий вибір обумовлений спрощеним візуальним діагностуванням.

За допомогою експертної оцінки було визначено два класи розпізнавання, які зустрічаються на рисунку 1.6: клас X_1^o – гіперхромні ядра; X_2^o – сполучна тканина.

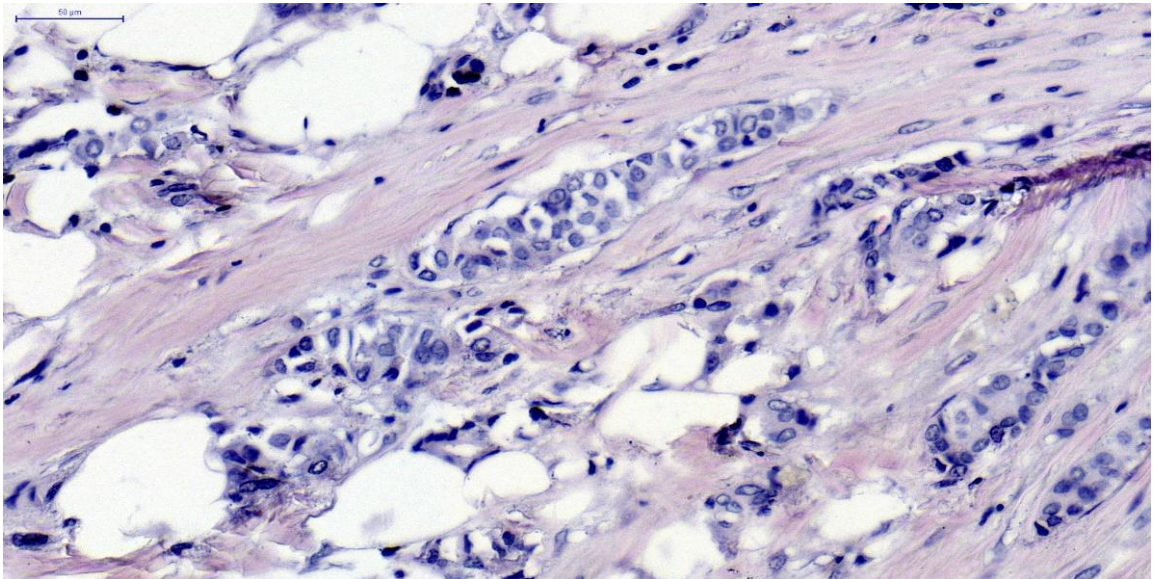


Рисунок 1.6 – знімок тканини, яка вражена злоякісними новоутвореннями

2 ОПИС МЕТОДУ ДОСЛІДЖЕННЯ

2.1 Формалізація алгоритму сегментації тканини

Невід’ємним технічним етапом при роботі з знімками отриманими в ході цифрової мікроскопії є сегментація вхідного зображення. При цьому, для задач локалізації ураження в тканині алгоритмів сегментації цілком достатньо, адже патологія, зазвичай, добре контрастує зі здоровими тканинами. Тому у деяких випадках, навіть без інтелектуальної компоненти можна чітко локалізувати осередок раку. У деякому сенсі, сегментація потрібна для нормалізації та спрощення вхідного зображення. Проте, на практиці сегментоване зображення додатково верифікується в деякій інтелектуальній компоненті.

Перший спосіб сегментації – це метод Оцу [12], який полягає у встановленні порогового значення для растрового знімку шляхом його бінаризації. Важливо, щоб вхідне зображення було у відтінках сірого. Ідея алгоритму полягає у порівнянні зважених сум усіх пікселів, щоб виділити найбільш контрастні від фону. Поріг яскравості (2.1) обчислюється шляхом поділу двох класів таким чином, щоб комбінований розкид був мінімальним.

$$S_w^2(x) = w_1(x)S_1^2(x) + w_2(x)S_2^2(x), \text{ де} \quad (2.1)$$

w_1 – вірогідність, що піксель відноситься до фону

w_2 – вірогідність, що піксель не відноситься до фону

x – порогове значення, що розділяє вірогідності двох класів

S_1 – дисперсія класу 1

S_2 – дисперсія класу 2

Наскільки би добре не була виконана початкова бінаризація зображення, завжди залишиться шум. Варто пам’ятати, що на гістологічних знімках найбільш контрастні ділянки – це ураження тканини. При цьому вони мають формувати чітко окреслені осередки. Це означає, що малоімовірна поява ракових залоз розміром у декілька десятків пікселів далеко від центру ураження.

Якщо припустити, що чорні пікселі на бінаризованому зображенні – це фон, а білі – це зона інтересу, то видалення зашумлення відбуватиметься за наступним алгоритмом:

- 1) Розширити зону інтересу шляхом збільшення світлих ділянок.
- 2) Розрахунок відстані між кожним пікселем, що характеризує зону інтересу та фоном.
- 3) Видалити усі надлишкові пікселі шляхом маніпуляції з відстанями (2).

Припускається, що усі вхідні зображення будуть зроблені в приблизно однакових умовах, тому параметри фільтрації та початкової сегментації можна зробити статичними.

Після отримання попередньо бінаризованого зображення настає етап безпосередньої сегментації. На цьому етапі варто розуміти, що для виконання цієї задачі не має єдино правильного, універсального рішення. Ефективність алгоритмів сегментації відрізняється в залежності від вхідного зображення. На практиці, вибір того чи іншого методу обумовлений співвідношенням точності результату до трудомісткості обчислень.

Одним з способів сегментації зображення – є метод встановлення рівня [15]. У рамках якого вхідне зображення розглядається як проекція деякої трьох вимірної площини, яка повинна як найкраще описати зону інтересу.

$$\frac{\partial \phi}{\partial t} = v |\nabla \phi|, \text{ де} \quad (2.2)$$

ϕ – це показник функції набору рівня;

$\nabla \phi$ – крок зміни рівня визначеної функції;

v – швидкість зміни контуру сегментованого зображення;

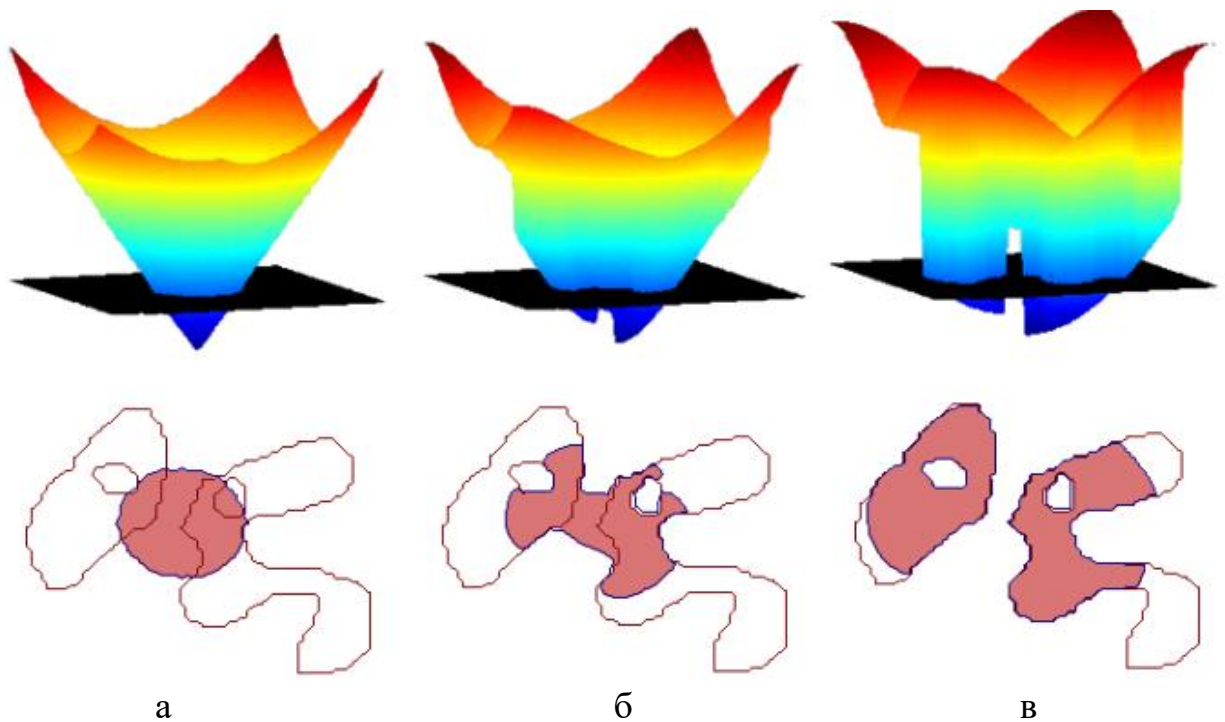


Рисунок 2.1 – Приклад використання методу встановлення рівня: а) початковий контур; б) проміжний результат розділення контуру; в) формування двох нових контурів

Візуально процес пошуку оптимальних контурів за методом встановлення рівнів можна описати як заповнення пористої поверхню рідиною. Цей підхід є не типовим для обробки зображення та комп'ютерного зору, адже існують набагато простіші способи визначити контури об'єкту інтересу. Тим не менш, саме цей метод набув широкого поширення при сегментації радіологічних зображень. Зокрема і для сегментації тканин молочної залози.

Також для знаходження контурів знімків використовують алгоритм К-середніх. Для цього припускається, що вхідне бінаризоване зображення розбите на кластери кожен із яких представлений як центроїд. Мета алгоритму полягає в мінімізації суми квадратів відстаней між центрами кластерів (2.3).

$$Q = \sum_{a=1}^k \sum_{b=1}^n \|x_b^a - C_a\|^2, \text{ де} \quad (2.3)$$

C_a – центр кластеру а;

x_b^a – положення поточного пікселю;

n – загальна кількість пікселів на зображенні;

Метод К-середніх є надзвичайно поширеним прийомом для вирішення прикладних задач в сфері машинного навчання. Тим не менш, у випадку сегментації цей алгоритм чутливий до шумів. Тому перед його використанням необхідно додатково упевнитися, що вхідне зображення було адекватно бінаризовано.

Одним з найбільш розповсюджених методів сегментації зображення є алгоритм водорозділу. Припускається, що будь-яке зображення у відтінках сірого можна розглядати як топографічну поверхню, де висока інтенсивність позначає вершини і пагорби, а низька визначає долини. Якщо призначити кожному локальному мінімуму певну мітку, то при зміщенні мінімального екстремуму він почне охоплювати все більше навколишніх ділянок. Це і є контури поточного об'єкту інтересу.

Тим не менш цей підхід дає надмірно сегментований результат через шум або інші нерівності зображення. Таким чином, наприклад, в OpenCV реалізовано алгоритм вододілу на основі маркерів, в якому користувач вказує, які точки западини повинні бути об'єднані, а які ні. Це інтерактивна сегментація зображень.

Один з варіантів цього алгоритму можна описати наступним чином:

- 1) Обирається набір маркерів, пікселів, з яких має початися «заливання». Кожному надається окрема мітка.
- 2) Сусідні пікселі кожної позначеної області переносяться в пріоритетну чергу, де пріоритет відповідає величині градієнта пікселя.
- 3) Піксель із найнижчим рівнем пріоритету витягується з черги. Якщо всі сусіди виділеного пікселя, які вже були позначені, мають однакову мітку, тоді піксель буде позначено їх міткою. Усі непозначені сусіди, які ще не перебувають у черзі пріоритетів, поміщаються до черги пріоритетів.

4) Повторіть крок 3, доки пріоритетна черга не буде порожньою.

Алгоритм водорозділу показує більшу ефективність, ніж інші методи сегментації, оскільки більше нагадує принцип роботи людського ока. Цей метод набув широкого поширення в додатках медичної візуалізації та комп'ютерного зору і є важливим кроком у багатьох комплексних системах обробки зображень. Незважаючи на свої обмеження, алгоритм водорозділу залишається популярним вибором для завдання сегментації зображень із-за його можливості обробляти знімки зі значною кількістю шумів і неправильної форми.

Для задачі локалізування ураження тканини алгоритми сегментації зображення – є найкращим вибором, адже не потребують значних обчислювальних навантажень. При цьому, гістологічні знімки характеризуються високою контрастністю зони інтересу, що спрощує їх сегментування.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Короткий опис програмної реалізації

Для реалізації програмного застосунку було використано мову програмування Python. Вибір обумовлений тим фактом, що на цій мові реалізовано велика кількість фреймворків пов'язаних з машинним навчанням, комп'ютерним зором та розпізнаванням образів.

Для роботи було використано наступні Python бібліотеки:

OpenCV для обробки та взаємодії з зображеннями. Зокрема, були задіяні морфологічні операції, такі як: дилатація та ерозія для мінімізації зашумлення зображення. Також були використанні вбудовані методи бінаризації зображення та дистанційні алгоритми сегментації.

Numpy для конвертації растрового зображення в математичний об'єкт для спрощення подальшої взаємодії з ним.

3.2 Результати моделювання

Як вхідне зображення розглядався гістологічний знімок молочної залози, рисунок 1.6, який містить наступні класи розпізнавання, рисунок 3.1, де X_1^o – це гіперхромія ядер залози; X_2^o – це сполучна тканина;

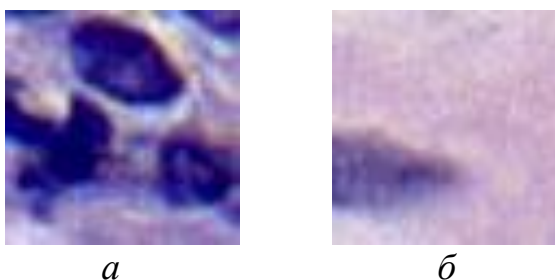


Рисунок 3.1 – Зображення кадрів зон інтересу: a – клас X_1^o ; b – клас X_2^o ;

Із-за специфіки обробки біоптату на знімку молочної залози (рис.1.6) онкологічні новоутворення контрастно відрізняються від сполучної тканини. Таким чином, їх можна візуально розділити навіть без використання додаткових

методів обробки зображення. Тим не менш, варто пам'ятати, що наданий знімок (рис. 1.6) – це в десятки разів збільшений фрагмент повнослайдового зображення.

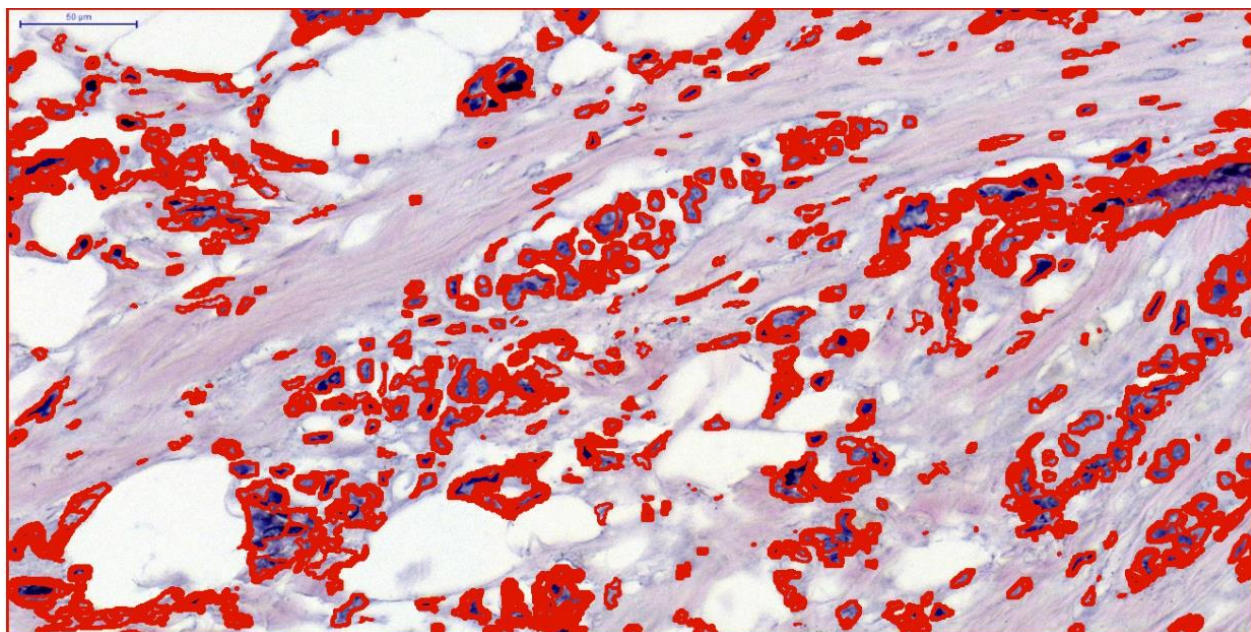


Рисунок 3.2. Сегментоване гістологічне зображення

Після описаного у другому розділі алгоритму сегментації, який включає в себе бінаризацію вхідного зображення з подальшим пошуком контурів зони інтересу на ньому, було отримано наступний результат, рисунок 3.2. Його можна вважати високо достовірним, адже зону інтересу, у вигляді гіперхромії клітин, було правильно локалізовано.

У подальшому необхідно додати до вже розробленої системи підтримки прийняття рішень інтелектуальну компоненту, яка б дозволила верифікувати результат сегментації. Таким чином, був би створений додатковий квантор перевірки ефективності алгоритму розпізнавати злоякісні новоутворення. Оскільки, поточна реалізація СППР може прийняти за гіперхромію ядер іншу ознаку тканини.

Важливим покращенням стане розширення алфавіту класів розпізнавання. У такому випадку задача сегментація доповниться вимогою класифікації, адже у поточній реалізації передбачається, що на зображенні знаходиться лише сполучна тканина або злоякісне новоутворення.

2. ВИСНОВКИ

Було розроблено алгоритмічний та програмний апарат, який сегментує вхідне гістологічне зображення на два найбільш релевантних класу розпізнавання: здорова та вражена онкологією тканина молочної залози. Використанні методи дозволяють системі буди максимально ефективною при мінімальному використанні обчислювальних можливостей.

Під час дослідження було проаналізовано способи діагностування раку молочної залози, які використовують у своїй практиці більшість лікарів онкологів. Отримана інформація дозволила подолати ряд програмних обмежень при розробці алгоритму. Зокрема, стали у нагоді умови використання цифрової мікроскопії в медицині.

3. СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

4. Breast cancer. <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed 13.05.2024)
5. Breast cancer Ukraine. <https://gco.iarc.fr/today/data/factsheets/populations/804-ukraine-fact-sheets.pdf> (accessed 13.05.2024)
6. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2017. – Vol. 39, No. 12. – P. 2481-2495.
7. Rozhkova N.I., Roitberg P.G., Varfolomeeva A.A. Mazo M.M., Dobrenkii A.N., Blinov D.S., Sushkov E.V., Deryabina O.N., Sokolov A.I. Neural network-based segmentation model for breast cancer X-ray screening. Sechenov Medical Journal. 2020; 11(3): 4–14. <https://doi.org/10.47093/2218-7332.2020.11.3.4-14>
8. Luo, W. Use cases for digital pathology / W. Luo, L. A. Hassell // Digital Pathology: Historical Perspectives, Current Concepts and Future Applications. – Springer Intern. Publ., 2016. – P. 5–15.
9. Beckwith, B. A. Standards for digital pathology and whole slide imaging / B. A. Beckwith // Digital Pathology: Historical Perspectives, Current Concepts and Future Applications. – Springer Intern. Publ., 2016. – P. 87–97.
10. Rozhkova N.I., Roitberg P.G., Varfolomeeva A.A. Mazo M.M., Dobrenkii A.N., Blinov D.S., Sushkov E.V., Deryabina O.N., Sokolov A.I. Neural network-based segmentation model for breast cancer X-ray screening. Sechenov Medical Journal. 2020; 11(3): 4–14. <https://doi.org/10.47093/2218-7332.2020.11.3.4-14>
11. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention. – Springer, Cham, 2015. – P. 234-241

- 12.Савченко, Т. Р. Інформаційна технологія персоналізованого діагностування ранньої стадії раку передміхурової залози. MS thesis. Сумський державний університет, 2023.
- 13.McKinney, Scott Mayer, et al. "International evaluation of an AI system for breast cancer screening." *Nature* 577.7788 (2020): 89-94.
- 14.Ye, Jiandong, et al. "Breast cancer image classification on WSI with spatial correlations." *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019.
- 15.Dovbysh, A. S., et al. "Decision-Making support system for diagnosis of breast oncopathologies by histological images." *Cybernetics and systems analysis* 59.3 (2023): 493-502.
- 16.Fan, Kun, Shibo Wen, and Zhuofu Deng. "Deep learning for detecting breast cancer metastases on WSI." *Innovation in Medicine and Healthcare Systems, and Multimedia: Proceedings of KES-InMed-19 and KES-IIMSS-19 Conferences*. Springer Singapore, 2019.
- 17.Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." *Automatica* 11.285-296 (1975): 23-27.
- 18.Houssami, Nehmat, et al. "Artificial intelligence for breast cancer screening: opportunity or hype?." *The Breast* 36 (2017): 31-33.
- 19.Dovbysh, Anatoliy, et al. "Decision-making support system for diagnosis of oncopathologies by histological images." *Journal of Pathology Informatics* 14 (2023): 100193.
- 20.Amarapur B (2020) Computer-aided diagnosis applied to MRI images of brain tumor using cognition based modified level set and optimized ANN classifier. *Multimedia Tools Appl* 79(5):3571–3599
- 21.Naji, Mohammed Amine, et al. "Machine learning algorithms for breast cancer prediction and diagnosis." *Procedia Computer Science* 191 (2021): 487-492.
- 22.Довбиш А.С. Основи проектування інтелектуальних систем: Навчальний посібник / А.С. Довбиш.– Суми: Видавництво СумДУ, 2009.– 172 с.

23. Dovbysh, A., et al. "The application of artificial intelligence in the diagnosis of prostate cancer." VIRCHOWS ARCHIV. Vol. 481. No. SUPPL 1. ONE NEW YORK PLAZA, SUITE 4600, NEW YORK, NY, UNITED STATES: SPRINGER, 2022.
24. Довбиш А.С. Інтелектуальні інформаційні технології в електронному навчанні / А.С. Довбиш, А.В. Васильєв, В.О. Любчак. – Суми: Видавництво СумДУ, 2013.– 177 с.
25. Chun-Fu Lin, Sheng-De Wang. Fuzzy support vector machines. IEEE Transactions on Neural Networks, 2002. pp 464-471. doi: 10.1109/72.991432
26. S.V.M. Vishwanathan, M. Narashimha Murty. SSVM: a simple SVM algorithm. IEEE 2002. doi: 10.1109/IJCNN.2002.1007516.
27. UCI Machine Learning Repository: Datasets. Access: <http://archive.ics.uci.edu/ml/machine-learningdatabases/breast-cancer-wisconsin/>

28. ДОДАТОК

```
import cv2
import numpy as np
import os

import matplotlib.pyplot as plt
from IPython.display import Image, display

def imshow(img, ax=None):
    if ax is None:
        ret, encoded = cv2.imencode(".jpg", img)
        display(Image(encoded))
    else:
        ax.imshow(cv2.cvtColor(img, cv2.COLOR_BGR2RGB))
        ax.axis('off')

img = cv2.imread("tumor.JPG")
imshow(img)

gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
imshow(gray)

ret, bin_img = cv2.threshold(gray,
                              0, 255,
                              cv2.THRESH_BINARY_INV +
cv2.THRESH_OTSU)
imshow(bin_img)

bin_img = cv2.bitwise_not(bin_img)
imshow(bin_img)

kernel = cv2.getStructuringElement(cv2.MORPH_RECT, (3,
3))
bin_img = cv2.morphologyEx(bin_img,
                           cv2.MORPH_OPEN,
                           kernel,
                           iterations=2)

imshow(bin_img)

fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(8,
8))
sure_bg = cv2.dilate(bin_img, kernel, iterations=5)
imshow(sure_bg, axes[0,0])
```

```

axes[0, 0].set_title('Sure Background')

dist = cv2.distanceTransform(bin_img, cv2.DIST_L2, 5)
imshow(dist, axes[0,1])
axes[0, 1].set_title('Distance Transform')

ret, sure_fg = cv2.threshold(dist, 0.2 * dist.max(), 255,
cv2.THRESH_BINARY)
sure_fg = sure_fg.astype(np.uint8)
imshow(sure_fg, axes[1,0])
axes[1, 0].set_title('Sure Foreground')

unknown = cv2.subtract(sure_bg, sure_fg)
imshow(unknown, axes[1,1])
axes[1, 1].set_title('Unknown')

plt.show()

ret, markers = cv2.connectedComponents(sure_fg)

markers += 1
markers[unknown == 255] = 0

fig, ax = plt.subplots(figsize=(6, 6))
ax.imshow(markers, cmap="tab20b")
ax.axis('off')
plt.show()

markers = cv2.watershed(img, markers)

fig, ax = plt.subplots(figsize=(5, 5))
ax.imshow(markers, cmap="tab20b")
ax.axis('off')
plt.show()

labels = np.unique(markers)

coins = []
for label in labels[2:]:

# Create a binary image in which only the area of the
label is in the foreground
#and the rest of the image is in the background

```

```
target = np.where(markers == label, 255,
0).astype(np.uint8)

# Perform contour extraction on the created binary
image
contours, hierarchy = cv2.findContours(
    target, cv2.RETR_EXTERNAL,
cv2.CHAIN_APPROX_SIMPLE
)
coins.append(contours[0])

# Draw the outline
img = cv2.drawContours(img, coins, -1, color=(0, 23,
223), thickness=2)
imshow(img)
```