



Współfinansowane ze
środków Unii Europejskiej



DOI: 10.5604/20830157.1148052

ZASTOSOWANIE I OGRANICZENIA TECHNOLOGII INFORMACYJNYCH W DIAGNOSTYCE MEDYCZNEJ

Marcin Maciejewski

Politechnika Lubelska, Instytut Elektroniki i Technik Informatycznych

Streszczenie. Artykuł zawiera przegląd zastosowań nowoczesnej technologii informacyjnej w medycynie i diagnostyce oraz wskazuje trudności na drodze postępującej informatyzacji wyżej wymienionych dziedzin. Do najczęściej stosowanych narzędzi informatycznych w służbie zdrowia należą bazy danych, narzędzia bazujące na algorytmach decyzyjnych i metodach przetwarzania danych. Do najważniejszych przeszkód w stosowaniu w/w metod zalicza się heterogeniczność danych medycznych, ich złożoność i konieczność interpretacji opisów słownych.

Słowa kluczowe: dane medyczne, przetwarzanie danych, heterogeniczność, data mining

ЗАСТОСУВАННЯ ТА ОБМЕЖЕННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ В МЕДИЧНІЙ ДІАГНОСТИЦІ

Анотація. Стаття містить огляд інформації про застосування сучасних інформаційних технологій у медицині та діагностиці, а також вказуються труднощі на шляху комп'ютеризації в вищезазначених областях. До найчастіше використовуваних інструментів у галузі охорони здоров'я відносяться бази даних, засоби які опираються на алгоритмах прийняття рішень і методах обробки даних. Найважливішими перешкодами під час застосування вищезазначених методів є багатозначність медичних даних, їх неоднорідність і необхідність інтерпретації вербальних описів.

Ключові слова: медичні дані, обробка даних, неоднорідність, data mining

INFORMATION TECHNOLOGY IMPLEMENTATIONS AND LIMITATIONS IN MEDICAL RESEARCH

Abstract. The article presents an overview of common uses of information technology in medicine and medical diagnostics, also pointing out major obstacles in the process of introducing information technology in the fields above. Information technology tools widely used in medicine include but are not limited to databases, decision algorithms and data processing and mining methods. Major obstacles include heterogeneity of medical data, their complexity and free text descriptions of procedures, diagnoses and interpretations of test results.

Keywords: medical data, data processing, heterogeneity, data mining

Wstęp

Dzięki niezliczonym odkryciom na polu nauk ścisłych jesteśmy obecnie w stanie dokonywać przetwarzania ogromnych ilości informacji z wykorzystaniem urządzeń mieszczących się na naszym biurku. Szczególnie takie dziedziny jak fizyka, inżynieria, matematyka i statystyka przyczyniają się do błyskawicznego rozwoju technologii informacyjnych poprzez poprawianie przepustowości łącz i miniaturyzację układów oraz zwiększanie ich mocy obliczeniowej. Jednocześnie, nasza cywilizacja dąży do ciągłego zwiększania długości życia poprzez ciągłe ulepszanie metod diagnostycznych i leczniczych, co nie mogłoby się odbyć bez intensywnych prac nad dokładnym zrozumieniem podstaw działania naszych organizmów. Jest to możliwe poprzez rozwój medycyny, biologii, chemii i biofizyki. Nie jest niczym nowym stwierdzenie, że obecnie nie jest możliwe prowadzenie zaawansowanych badań statystycznych bez znacznego wsparcia ze strony technologii informacyjnych, szczególnie w przypadkach kiedy przestrzeń danych jest zbyt rozległa lub skomplikowana do opracowania przez jednego człowieka. Kluczowymi są szczególnie aspekty gromadzenia, zarządzania i przetwarzania danych [11].

1. Magazynowanie i zarządzanie danymi medycznymi

Ciągły rozwój technologii diagnostycznych prowadzi do stałego powiększania się puli dostępnych dla lekarzy testów. W wyniku tego znacznie rośnie też ilość generowanych przez nie danych wynikowych. Aby można było mówić o jakiegokolwiek użyteczności tych danych, konieczne jest ich magazynowanie i zarządzanie nimi w sposób umożliwiający zainteresowanym kontrolowany dostęp [11]. W okresie poprzedzającym informatyzację służby zdrowia realizowane było to poprzez budowanie

Вступ

Завдяки численним відкриттям в галузі науки, сьогодні ми в стані виконувати обробку величезних обсягів інформації використовуючи при цьому пристрої які знаходяться на нашому столі. Особливо в таких галузях, як фізика, техніка, математика і статистика. Інформаційні технології швидко розвиваються за рахунок поліпшення використання смуги пропускання каналів і мініатюризації систем зі збільшенням їх потужності. У той же час, наша цивілізація прагне до безперервного зростання тривалості життя за рахунок постійного вдосконалення методів діагностики та лікування, які не могли б здійснюватися без точного розуміння принципів роботи наших органів. Це стало можливим завдяки розвитку медицини, біології, хімії та біофізики. Не є новиною той факт, що зараз не можливо провести додаткові статистичні обстеження без істотної підтримки з боку інформаційних технологій, особливо в тих випадках, коли кількість даних занадто велика і складна, щоб опрацювала одна людина. Ключовими є аспекти нагромадження, управління та обробки даних [11].

1. Управління та зберігання медичних даних

Систематичний розвиток діагностичних технологій призводить до постійного розширення кількості доступної інформації для лікарів. В результаті, значно збільшується, кількість згенерованих ними даних. Щоб можна було говорити про будь-яку користь цих даних, необхідним є їх зберігання і управління таким чином, щоб дозволити зацікавленим контролюваний доступ [11]. Коли охорона здоров'я ще не була комп'ютеризована зберігання інформації реалізовувалося шляхом створення паперових архівів, якими

papierowych archiwów zarządzanych przez odpowiednio przeszkolonych pracowników. Obecnie, takie podejście zostało zastąpione przez wyspecjalizowane do określonego zadania komputerowe bazy danych ze zdalnym dostępem.

1.1. Medyczne bazy danych

Medyczne bazy danych mogą zostać podzielone na szereg kategorii [7]:

- Kliniczne bazy danych, zawierające głównie znaczne ilości danych pacjentów i wyników ich badań koniecznych do diagnostyki. Mogą one być stosowane jako dane wejściowe zarówno do analiz opisowych jak i predykcyjnych.
- Bazy danych pacjentów, zawierające dane dla określonego pacjenta, budowane podczas leczenia oraz opieki szpitalnej. Mogą nierzadko udostępniać zawartość wielu użytkownikom, a ich budowa może się dokonywać przez całe życie pacjenta.
- Specjalistyczne kliniczne bazy danych. Mogą zawierać informacje specyficzne dla danego schorzenia, urzędnika, procedury lub populacji. Wykorzystywane są do gromadzenia rekordów z wielu klinicznych baz danych.
- Bazy danych badań klinicznych wykorzystywane podczas badań klinicznych, takich jak monitorowanie skuteczności leków czy procedur.
- Bazy danych nadzoru biologicznego, służące do pomocy podczas analizowania skutków ubocznych stosowania leków oraz epidemii.
- Bazy metadanych, służące jako swoiste słowniki dla innych baz danych.

Podczas implementacji oraz stosowania medycznych baz danych konieczne jest stosowanie się do odpowiednich wymagań funkcjonalnych.

1.2. Zarządzanie danymi medycznymi i ich standaryzacja

W celu poprawienia jakości pracy w systemach medycznych konieczne jest wprowadzenie szeregow wytycznych narzucających odgórnie pewne standardy [1]. Z uwagi na znaczną ilość przedsięwzięć zajmujących się produkcją oprogramowania i sprzętu medycznego konieczne jest zapewnienie pewnego poziomu kompatybilności skrośnej i wstecznej. Problem standaryzacji danych jest bardzo istotny w służbie zdrowia z uwagi na liczne przypadki braku kompatybilności wzajemnej licznych baz danych znajdujących się nawet na jednym oddziale szpitala i konieczności ręcznego przepisywania danych pomiędzy nimi. Jest to przyczyną wielu błędów i nieprawidłowości z uwagi na prawdopodobieństwo wystąpienia pomyłek operatora.

Prawdopodobnie najbardziej rozpowszechnionym standardem jest The Clinical Data Interchange Standards Consortium (CDISC) [16]. Jest on tworzony od 1997 w ramach organizacji non-profit. Jej celem jest stworzenie i rozwijanie globalnego, niezależnego od platformy i implementacji standardu wymiany danych pozwalającego na usprawnienie procedur badawczych i wielu obszarów służby zdrowia (rys. 1).

W podobnym celu stworzony został również standard opracowany przez organizację HL7 (Health Level Seven) [18]. Głównym celem organizacji jest wprowadzenia zasad kierujących wymianą, integracją, współdzieleniem i zdobywaniem elektronicznej informacji medycznej w praktyce klinicznej oraz zarządzaniu usługami w służbie zdrowia (rys. 2). Sama nazwa wywodzi się z sieciowego modelu warstw komunikacji OSI i oznacza integrację na siódmej warstwie – warstwie aplikacji.

Inną obecnie rozwijającą się organizacją jest openEHR (open Electronical Health Record) [20], stawiającą sobie za cel zbudowanie języka służącego do budowy modeli klinicznych lub archetypów, niezależnych od platformy, bazy i zastosowania (rys. 3).

керували кваліфіковані працівники. В даний час цей підхід замінений на спеціалізовані для виконання конкретного завдання комп'ютерні бази даних з віддаленим доступом.

1.1. Медичні бази даних

Медичні бази даних можуть бути розділені на кілька категорій [7]:

- Поліклінічні бази даних, які містять в основному великі обсяги даних пацієнтів і результати їх обстежень, необхідні для встановлення діагнозу. Вони можуть бути використані в якості вихідних даних для описового аналізу та прогнозування
- Бази даних пацієнтів, в яких знаходяться дані для конкретного пацієнта під час лікування в медичному закладі та для опіки після виписки з лікарні. До їх інформації часто мають доступ кілька користувачів, а також їх формування може відбуватися протягом усього життя пацієнта.
- Спеціалістичні клінічні бази даних. Можуть містити інформацію, яка відноситься до конкретної хвороби, пристроїв, процедур або населення. Вони використовуються для збору результатів з декількох клінічних баз даних.
- Бази даних, використовуваних в лікарняних дослідженнях, такі як контроль ефективності лікарських препаратів або процедур.
- Бази даних біомоніторингу, які допомагають при аналізі побічних ефектів лікування та епідемії
- Метадані бази даних, які використовують, свого роду, словники для інших баз даних.

Під час впровадження та використання медичних баз даних необхідно притримуватися відповідних функціональних вимог.

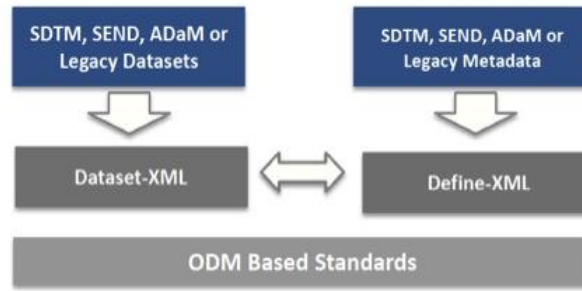
1.2. Управління медичними даними і їх стандартизація

З метою підвищення якості роботи в медичних системах необхідно дотримуватися ряду стандартів [1]. Враховуючи велику кількість компаній, що беруть участь у виробництві програмного забезпечення і медичної техніки необхідно забезпечити певний рівень перехресної сумісності. Проблема стандартизації даних важлива в охороні здоров'я через численні випадки взаємної несумісності багатьох баз даних, які знаходяться навіть в одному відділенні лікарні, а також необхідність вручну переписувати дані між ними. Це причина багатьох помилок і порушень у зв'язку з імовірністю помилки оператора.

Найбільш розповсюдженим стандартом є The Clinical Data Interchange Standards Consortium (CDISC) [16]. Він створений в 1997 році в рамках організації non-profit. Її метою є створення і розвиток глобального, незалежного від платформи і реалізації стандарту обміну даними для спрощення процедури досліджень в багатьох галузях охорони здоров'я (мал. 1).

З подібною метою був також створений стандарт, розроблений організацією HL7 (Health Level Seven) [18]. Основною метою організації є впровадження принципів, що регулюють обмін, інтеграцію, спільне використання та набуванням електронної медичної інформації в клінічній практиці і управління послугами в галузі охорони здоров'я (мал. 2). Сама назва походить від мережевої моделі рівнів зв'язку OSI і означає інтеграцію на сьомому шарі – це додатковий шар.

Іншою організацією яка розвивається в даний час є openEHR (open Electronical Health Record) [20], метою якої є створення мови для побудови клінічних моделей або архетипів, незалежних від платформи, бази і застосування (мал. 3).

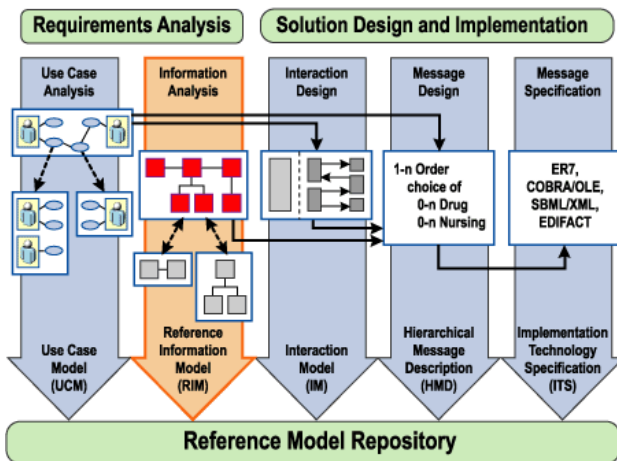


Rys. 1. Schemat wymiany danych i komunikatów w systemie CDIS [17]

Мал. 1. Схема обміну даними в системі CDIS [17]

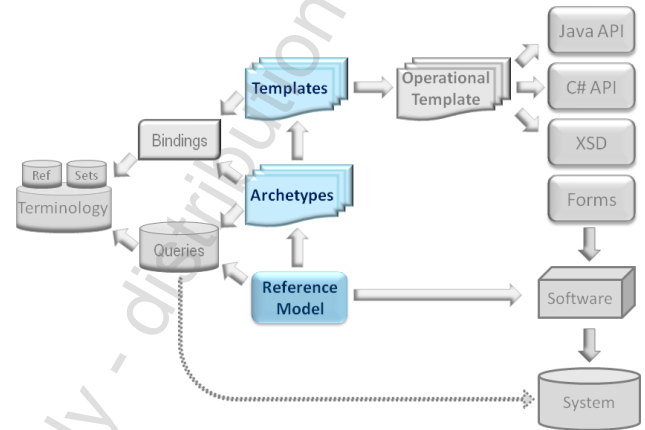
The HL7 Message Development Framework

Phrases, activities, and models



Rys. 2. Budowanie systemu wymiany informacji w HL7 [19]

Мал. 2. Будова системи обміну інформації в HL7 [19]



Rys. 3. Idea niezależnych modeli openEHR [21]

Мал. 3. Ідея незалежних моделей openEHR [21]

1.3. Dane medyczne

Poprzez pojęcie danej medycznej rozumiemy wynik jednej obserwacji stanu pacjenta, na przykład temperatury, ciśnienia lub wycinka historii chorób. Najczęściej, dana medyczna składa się z czterech określonych pól: identyfikatora pacjenta, badanej wielkości, wartości wielkości oraz daty badania [1].

Jednakże, pojęcie danych medycznych ma znacznie szersze znaczenie. Można wyróżnić pewne określone rodzaje danych medycznych pod względem semantycznym, między innymi [13]:

- kategorie, operujące na zbiorach nieuporządkowanych, na przykład grupa krwi {0+, A+, B+ ...},
- kategorie uporządkowane, na przykład {niski, średni, wysoki},
- cykliczne, dla których zbiorów wartości zawiera powtarzający się, skończony zbiór wartości, na przykład minuty, godziny, dni tygodnia,
- hierarchiczne, operujące na usystematyzowanych właściwościach, na przykład oznaczenia jednostek chorobowych w kategoriach Międzynarodowej Klasyfikacji Chorób ICD-9 [22],
- grafy, zawierające elementy i relacje między nimi. Przykładem jest opis anatomiczny ludzkiego ciała lub lokalizacja ogniska zapalnego,
- zbiory danych, często będące podzbiórami skończonego zbioru zawierającego wszystkie zestawy wartości (skończony zbiór diagnoz). Jednakże, w niektórych przypadkach budowanie takiego zbioru nie jest praktyczne ze względu na znaczną przestrzeń danych,
- interwały,
- stosunki.

Innym problemem występującym podczas przetwarzania danych medycznych jest niejednokrotnie brak dostępu do pojedynczych danych dla każdego przypadku, a jedynie do

1.3. Медичні дані

Під поняттям медичні дані розуміємо результати спостережень пацієнта, наприклад його температури, тиску або історію хвороби. Найчастіше, медичні дані складаються з чотирьох конкретних областей: ID пацієнта, вибірка, розмір і дата дослідження [1].

Але поняття медичних даних, має набагато ширше значення. Можна виділити деякі специфічні види медичних даних, між іншим, з семантичної точки зору [13]:

- категорії, в основі яких є неструктуровані множини, наприклад групи крові {0+, A+, B+ ...},
- упорядковані категорії, наприклад {низький, середній, високий},
- циклічні, які містять повторюваний набір значень, такі як хвилини, години, дні тижня,
- ієрархічні, полягають на структурованих властивостях, наприклад, посилання в плані захворювань Міжнародній класифікації хвороб ICD-9 [22],
- графіки, що містять елементи і відносини між ними. Прикладом може бути опис анатомії людського тіла або місце розташування запалення,
- набори даних, які часто є підмножинами кінцевого набору, що містить всі набори значень (кінцевий збір діагнозів). Тим не менш, в деяких випадках, укладання такого набору не є практичним через велику кількість даних,
- інтервали,
- відносини.

Ще однією проблемою в обробці медичних даних є відсутність доступу до персональних даних для кожного окремого випадку, тільки до агрегації даних, наприклад, конкретна популяція з даної лікарні, або пацієнти, які мають

агрегacji danych np. danej populacji, z danej kliniki lub pacjentów cierpiących na jedno schorzenie. Nierzadko dane znajdujące się w postaci tabeli lub zestawienia zawierają metadane zastępujące brakujące lub niepewne wyniki badań, w postaci „?”, „brak”, „NA” lub innych.

2. Przetwarzanie danych medycznych

Konieczność analizy i wnioskowania na podstawie danych medycznych jest wywołana dążeniem do odkrycia nowych metod diagnostycznych i przesiewowych. Obecnie, dużym zainteresowaniem cieszą się systemy eksperckie, obiecujące wspomaganie podejmowania decyzji podczas diagnozy i leczenia choroby. Takie systemy wymagają odpowiedniego podejścia do problemu poprzez wykorzystanie odpowiedniej metodyki, co nie jest łatwe ze względu na szczególny charakter danych medycznych [8].

2.1. Heterogeniczność

W praktyce największym wyzwaniem dla twórców systemów zarządzania i przetwarzania danych medycznych jest ich heterogeniczność [2]. Mogą one składać się z danych obrazowych, opisowych, wywiadów z pacjentami, danych laboratoryjnych oraz spostrzeżeń i interpretacji lekarza. Wszystkie powyższe składowe powinny być brane pod uwagę podczas diagnozy, prognoz oraz leczenia pacjenta. Główne problemy wynikające z heterogeniczności danych medycznych można podzielić następująco:

- złożoność danych medycznych,
- konieczność uwzględnienia interpretacji lekarza,
- uwzględnienie czułości i specyficzności,
- niedokładny opis matematyczny,
- Forma zapisu danych.

Złożoność danych medycznych ma związek głównie ze znaczną ilością różnorodnych źródeł danych medycznych. Coraz więcej metod diagnostycznych pozwala na uzyskanie danych w formie obrazowej. Tomografia komputerowa, rezonans magnetyczny i inne badania obrazowe generują setki megabajtów danych dziennie, co powoduje konieczność budowy magazynów i baz danych z możliwościami wyszukiwania obrazowego. Ponadto, do obrazów są nierzadko dołączone interpretacje jednego lub wielu diagnostów oraz dane pacjenta. Rozpoznawanie wzorców w takich obrazach, chociaż intuicyjne dla człowieka, stanowi znaczne wyzwanie w automatyzacji procesu diagnostycznego.

Również przetworzenie interpretacji lekarza na język zrozumiały przez algorytm diagnostyczny nie jest zadaniem trywialnym. Wynika to z faktu, że interpretujący dane schorzenie lekarze mogą nie zgadzać się zarówno co do interpretacji danego schorzenia oraz używać różnego żargonu medycznego. Obecnie stosowane systemy rozpoznawania treści oferują znikomą dokładność w zdaniach zawierających powyżej 10 słów.

Niedokładny opis matematyczny danych medycznych wynika z charakteru medycyny jako nauki. Opis takich zjawisk jak zapalenie, zakrzep czy opuchlizna wymyka się twardym regułom formalizmu stosowanego w matematyce i fizyce. Lekarz i diagnosta nie posiada dokładnie zdefiniowanego modelu ani struktury w które mógłby wpasować swoje spostrzeżenia ani diagnozę. W chwili obecnej duże nadzieje wiąże się w stosowaniu logiki rozmytej i metod z dziedziny kopalni wiedzy.

Forma zapisu danych wiąże się ściśle z budową gramatyczną zdań składających się na diagnozę. Wyrażenia takie jak „zapalenie wątroby” czy „wątroba z widocznym zapaleniem” oznaczają to samo, jednakże mogą zostać różnie zinterpretowane przez automatyczny proces diagnostyczny. Jeszcze trudniejsze w interpretacji są wyrażenia opisujące ilość i stopień, na przykład „niektóre”, „wiele”, „znacznie”.

odnakową chorobę. Nierідкі випадки, коли дані у вигляді таблиці або списку, що містить метадані які замінюють відсутні або ненадійні результати випробувань, у формі “?”, “відсутні”, “NA” або інші.

2. Обробка медичних даних

Необхідність аналізувати і робити висновки на підставі медичних даних викликана бажанням відкрити нові діагностичні методи. В даний час великою популярністю користуються експертні системи, обіцяючи підтримку процесу прийняття рішень в діагностиці та лікуванні захворювань. Такі системи вимагають відповідного підходу до вирішення проблеми шляхом використання відповідної методології, а це складно через специфічний характеру медичних даних [8].

2.1. Гетерогенність

На практиці, великою проблемою для розробників систем управління та обробки медичних даних є їх неоднорідність [2]. Вони можуть складатися з даних візуальних, описових, інтерв'ю з пацієнтами, лабораторних даних і спостережень та інтерпретації лікаря. Всі ці компоненти повинні бути прийняті до уваги при діагностиці та прогнозуванню лікування пацієнта. Основні проблеми, пов'язані з неоднорідністю медичних даних можна поділити таким чином:

- багатогранність медичних даних,
- необхідність врахування інтерпретації лікаря,
- врахування специфіки,
- неточний математичний опис.
- форма запису даних.

Багатогранність медичних даних пов'язана, головним чином, зі значною кількістю різних джерел інформації про стан здоров'я. Все більше методів діагностики дозволяє отримати дані в формі малюнків. Комп'ютерна діагностиката, магнітичний резонанс і інші малюнкові методики дослідження генерують сотні мегабайт даних в день, що робить необхідним створення складів та баз даних з можливостями пошуку зображень. Крім того, зображення часто супроводжується одним або декількома інтерпретаціями діагнозів і даних пацієнта. Розпізнавання зразків у цих картинках, хоча б інтуїтивно для людини, являє собою значну проблему в автоматизації діагностичного процесу.

Також обробка інтерпретації лікаря на мову зрозумілу для діагностичного алгоритму не таке тривіальне завдання. Це пов'язано з тим, що описуючи хворобу, лікарі не можуть дійти згоди щодо інтерпретації даної хвороби і використовують різні медичні терміни. В даний час системи розпізнавання даних пропонують низьку точності в реченнях, що містять більше 10 слів.

Неточний математичний опис медичних даних виникає з походження медицини, як науки. Опис таких явищ, як запалення, тромбоз чи припухлість відходять від жорстких правил формальності, який використовується в математиці і фізиці. Лікар не має чітко визначеної моделі або структури, в яку міг би вписати свої спостереження або діагноз. В даний час великі надії покладають на застосування нечіткої логіки і методів у галузі пошуку знань.

Форма запису даних тісно пов'язана з формулюванням граматично правильних речень, при постановці діагнозу. Такі вирази, як “запалення печінки” або “печінка в якій спостерігається запалення”, означає те ж саме, однак, може бути зінтерпритовано по-різному через автоматизацію діагностичного процесу. Ще важче інтерпретувати вирази, що описують кількість і ступінь, наприклад “деякі”, “багато”, “більше”.

2.2. Metody przetwarzania danych

Ze względu na złożoność problemu analizy danych medycznych wykazaną w poprzedniej części pracy konieczne jest niejednokrotnie stosowanie metod ogólnych, koncentrujących się bardziej na opisie poszczególnych zbiorów przypadków niż na wyszukiwaniu specyficznych szczegółowych powiązań pomiędzy symptomami i chorobą. Taką grupę metod stanowią metody statystyczne, których łatwość stosowania i niska złożoność obliczeniowa sprawiły, iż są powszechnie stosowane przez badaczy medycznych na całym świecie [5]. Metody te najczęściej operują na grupach reprezentatywnych dla danej populacji, wyciągając na ich podstawie wnioski dotyczące całości.

Jednym z najstarszych sposobów prowadzenia wnioskowania, praktykowanym jeszcze w starożytności, jest kierowanie się zasadami logiki. Formalny zapis zależności pomiędzy zmiennymi poprzez wyrażenia logiczne typu I, LUB, JEŻELI ... TO przy wykorzystaniu teorii zbiorów pozwala na proste opisanie ich związków [12]. Wielokrotnie podkreślana heterogeniczność danych medycznych wynikająca ze wcześniej wymienionych czynników sprawia jednakże, iż rzadko możliwe jest wykorzystanie metod klasycznej logiki bez uprzedniego odpowiedniego przygotowania danych. Problemem jest tu najczęściej brak dokładnie zdefiniowanych wartości matematycznych dla nieprecyzyjnych opisów słownych, takich jak na przykład „lekki”, „ciężki”, „krytyczny”. Z tego powodu, aby korzystać z prostego języka logiki do opisu związków matematycznych w naukach przyrodniczych należy zastosować podejście logiki rozmytej [10]. Została ona po raz pierwszy wprowadzona do terminologii naukowej w 1965 roku [14]. Od tego czasu rozpowszechniła się wśród specjalistów z wielu dziedzin, począwszy od kontroli procesów, a na sztucznej inteligencji skończywszy. Jest to spowodowane wysoką elastycznością podejścia i wysoką stosowalnością podczas kontroli procesów i budowy systemów decyzyjnych.

Logika typu rozmytego zrywa z klasycznym podejściem prawda-fałsz podczas podejmowania decyzji. Zamiast tego proponuje wprowadzenie częściowo nakładających się przedziałów wartości wielkości wejściowych, co powoduje łatwość stosowania do przetwarzania informacji podanej w formie „mało-średnio-dużo”. Dzięki temu możliwe jest podejmowanie decyzji zarówno na podstawie danych liczbowych jak i na podstawie rozmytych w swojej postaci diagnoz lekarskich [9].

Medyczne bazy danych są też doskonałym materiałem źródłowym do stosowania obecnie szeroko rozwijanych metod z dziedziny data mining [3]. Idea kopalni danych zakłada poszukiwanie charakterystycznych cech w badanych zbiorach oraz zależności między nimi. Do wykorzystywanych algorytmów należą między innymi sieci neuronowe, drzewa decyzyjne oraz wszelkiego typu klasyfikatory.

Sztuczne sieci neuronowe zostały zaproponowane po raz pierwszy w 1943 roku jako matematyczny sposób opisu działania sieci neuronowych u żywych organizmów. Metoda korzysta z tak zwanej logiki progowej, opierającej się na założeniach zbliżonych do modeli regresyjnych i logiki rozmytej. Modele sieci wykorzystują szereg sztucznych, wielowarstwowych elementów decyzyjnych, najczęściej ułożonych w trzech warstwach, zwanych perceptronami lub węzłami (rys. 4).

Jedną z najszerszej stosowanych metod uczenia maszynowego z indukcją są drzewa decyzyjne [6]. Metoda stosuje odwzorowanie zbioru argument-wartość jako zestaw wielopoziomowych warunków logicznych (rys. 5). Taka forma zapisu jest intuicyjna dla człowieka i często stosowana w medycynie [4]. Metody bazujące na drzewach decyzyjnych są odporne na szum i łatwe w implementacji. Jako kryterium podziału w drzewach decyzyjnych jest wykorzystywana entropia informacji, przez co zawsze dąży się do najbardziej korzystnego podziału.

Klasyfikatory Bayesa operują na zasadzie minimalizacji prawdopodobieństwa niepoprawnej klasyfikacji przypadku poprzez założony model probabilistyczny [15]. Klasyfikacja może zostać dokonana poprzez porównywanie nowego przypadku z zestawem danych uczących lub ze zbudowanym na jego podstawie modelem.

2.2. Методи обробки даних

Через складність аналізу медичних даних зазначених у попередній частині статті часто доводиться використовувати загальні методи, які більше орієнтовані на опис кожного набору випадків, ніж на пошук конкретних детальних зв'язків між симптомами і хворобами. Такою групою є статистичні методи, вони прості у використанні і обчисленні, тому їх постійно використовують медичні дослідники по всьому світу [5]. Ці методи застосовують на вибірці представників однієї групи населення, а на основі отриманих результатів роблять висновки для всіх.

Один з найстаріших способів формування висновків, який практикувався ще в стародавні часи, є принцип логіки. Взаємозв'язок між змінними за допомогою логічного виразу типу I, АБО, ЯКЩО ... ТО за допомогою теорії множин забезпечує простий опис їхніх взаємовідносин [12]. Неодноразово наголошувалася на факт неоднорідності медичних даних який виникає в результаті вищезазначених факторів це призводить до того, що рідко можна використовувати методи класичної логіки без належної підготовки даних. Найчастішою проблемою є відсутність чітко визначених математичних значень для неточних словесних описів, таких як "легкий", "важкий", "критичний". З цієї причини, щоб використовувати просту мову для опису логіки математичних відносин в галузі природничих наук, використовується підхід неточної логіки [10]. Вона була вперше представлена в науковій термінології в 1965 році [14]. З того часу поширилася серед фахівців багатьох галузей, починаючи від контролю процесів і завершуючи штучним інтелектом. Це пов'язано з високою гнучкістю підходу і частим використанням в процесі контролю і будівництва процесів з прийняття рішень.

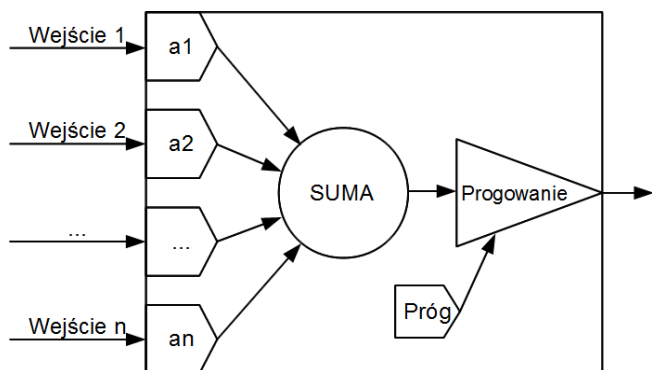
Нечітка логіка розходиться з класичним підходом правд-фальш при прийнятті рішень. Замість цього пропонує ввести межі вхідних змінних, які частково перекриваються, що робить його простим для обробки інформації, наведеної у вигляді "мало-трохи-багато". Це дозволяє приймати рішення, опираючись на кількісні дані, на основі нечітких медичних діагнозів [9].

Медичні бази даних також є відмінним вихідним матеріалом для широкого розповсюджених методів в області data mining [3]. Ідея інтелектуального аналізу даних включає в себе пошук подібних характеристик у досліджуваних і взаємозалежності між ними. До алгоритмів які використовуються належать нейронні мережі, дерева прийняття рішень, і всі типи класифікаторів.

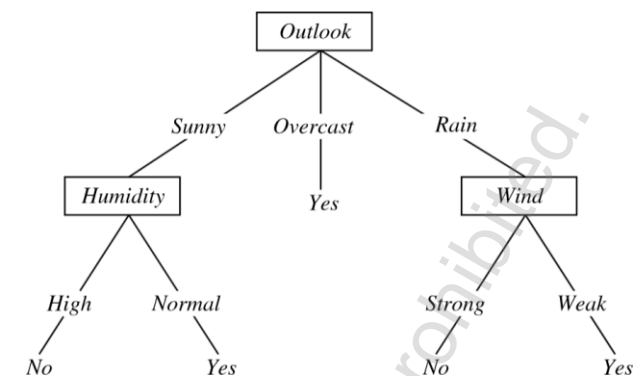
Штучні нейронні мережі були запропоновані вперше в 1943 році в якості математичного опису дій нейронних мереж в живих організмах. Метод використовує так звану порогову логіку, яка опирається на припущеннях приближених до регресійних моделей і нечіткої логіки. Мережеві моделі використовують ряд штучних, з кількома варіантами елементів прийняття рішень, які, як правило, вкладені в три шари, які називаються перцептронами або вузлами (мал. 4).

Одним з найбільш широко використовуваних методів машинного навчання з індукцією є дерево прийняття рішень [6]. Методика використовує відображення набору аргумент-вартість у вигляді набору багаторівневих логічних умов (мал.5). Така форма запису є інтуїтивною для людини і часто використовується в медицині [4]. Методики, які опираються на деревах прийняття рішень прості в імplementації. Як критерій для розподілу в деревах прийняття рішень використовується інформаційна ентропія, яка завжди скерована до найбільш корисного поділу.

Байєсівські класифікатори працюють за принципом мінімізації ймовірності неправильної класифікації випадку за допомогою встановленого пробабілістичного підходу [15]. Класифікація може бути зроблена шляхом порівняння нового випадку з набором навчальних даних або побудованим на його основі зразком.



Rys. 4. Schemat jednej komórki neuronowej z progowaniem
 Мал. 4. Схема однієї клітини нейронного порозу



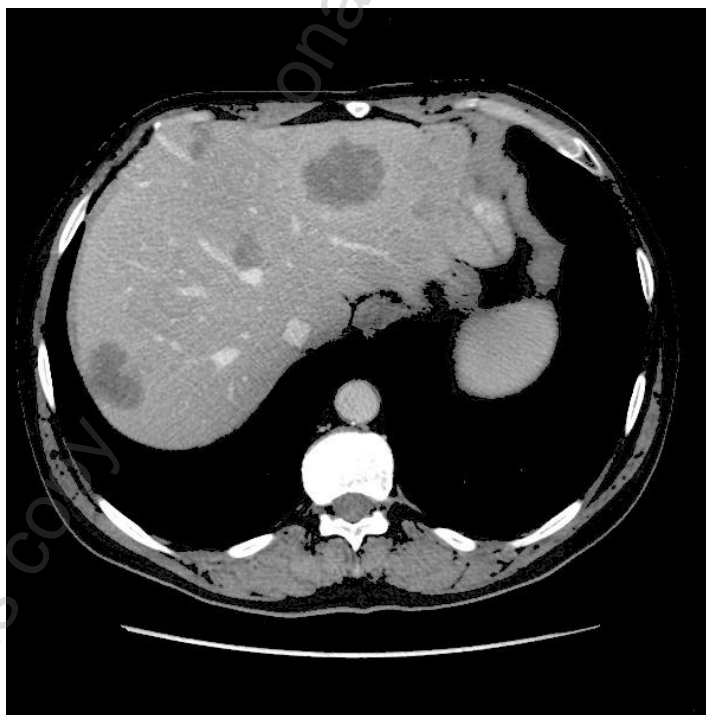
Rys. 5. Przykładowe drzewo decyzyjne [6]
 Мал. 5. Зразок дерева прийняття рішень [6]

2.3. Przetwarzanie danych obrazowych

Kompletnie niezależnym problemem jest kwestia przetwarzania obrazów medycznych. Obrazy takie niosą przeważnie ogromną ilość użytecznej podczas diagnozy i leczenia informacji. Konieczność włączenia ich jako danych wejściowych do procesu diagnostycznego powoduje dodatkową trudność polegającą na odpowiednim ich przetworzeniu. Jednakże, nawet najlepiej wykonany proces odszumiania i wykrywania krawędzi sam w sobie nie wystarczy do przeprowadzenia skutecznej diagnozy. Diagnostyka obrazowa w medycynie jest dziedziną wymagającą doświadczenia operatora oraz względu na cechy osobnicze oddzielnie dla każdego przypadku. Algorytm diagnostyczny wykorzystujący obrazy jako dane wejściowe musi być niezależny od tych zmian, co nierzadko powoduje ograniczoną przydatność metod obrazowania w niektórych przypadkach. Jak do tej pory niemożliwe jest kompletne zastąpienie specjalisty podczas dokonywania analizy i wystawiania diagnozy na podstawie obrazu, a programy implementujące szereg algorytmów przeważnie ograniczają się do poprawiania jakości i segmentacji obrazu, rzadko dokonując automatycznej i niewspomaganej analizy kształtu i tekstury. Przykładowy obraz pochodzący z tomografii komputerowej przedstawiony jest na rysunku 6 poniżej.

2.3. Обробка візуальних даних

Проблемою залишається питання обробки медичних малюнків. Ці зображення, як правило, несуть в собі величезну кількість корисної під час діагностики та лікування інформації. Необхідність включати їх в якості вступних даних до процесу діагностики викликає додаткове утруднення, що полягає на правильності їх обробки. Тим не менш, навіть краще зроблений процес зменшення шуму і виявлення країв сам по собі не є достатнім для ефективної діагностики. Візуальна діагностика в медицині є галузю, яка вимагає досвіду оператора, і врахування індивідуальних особливостей кожного конкретного випадку. Діагностичний алгоритм, що використовує в якості вхідних даних зображення, повинен бути незалежним від змін, які часто обмежують корисність методів візуалізації. До цих пір, неможливо повністю замінити фахівця під час аналізування та діагностування на основі образу, а програми імплементуючи ряд алгоритмів, зводяться до подальшого підвищення якості зображення та сегментації малюнка, рідко виконуючи автоматичний і без сторонньої допомоги аналіз форми і текстури. Прикладом може бути зображення з комп'ютерної томографії показаної на малюнку 6.



Rys. 6. Przykładowe obraz pochodzący z tomografii komputerowej na wysokości wątroby

Мал. 6. Приклад зображення з комп'ютерної томографії на рівні печінки

3. Podsumowanie

Rzeczywistość nowoczesnej medycyny nie może odbywać się bez odpowiedniego wsparcia ze strony komputerów oraz metod informatycznych. Przyczyniają się do tego takie czynniki jak znacząca wielkość baz danych, rozwój metod diagnostycznych, wzrost znaczenia badań obrazowych przeprowadzanych różnymi metodami i konieczność analizy wciąż rosnących ilości informacji. Poprzez zastosowanie odpowiednio przygotowanych systemów informatycznych oraz metod diagnostycznych możliwe jest ułatwienie pracy lekarzy oraz personelu medycznego. Jednakże, skuteczne wprowadzenie metod automatycznego wspomaganego diagnostyki jest szczególnie utrudnione w przypadku medycyny z powodu braku ścisłych, jednoznacznych metod opisów schorzeń i objawów które mogą być łatwo przetwarzane i wykorzystywane jako wejścia w systemach wspomaganego decyzyjnego. Daje to szerokie pole do popisu dla twórców metod z dziedziny data mining oraz w specjalistów od systemów decyzyjnych, uczenia maszynowego czy przetwarzania danych wielowymiarowych.

Podziękowania

Praca powstała w ramach projektu *PL-NTU Transgraniczna wymiana doświadczeń* PBU.03.01.00-06-386/11-00 współfinansowanego w ramach Programu Współpracy Transgranicznej Polska-Białoruś-Ukraina 2007-2013 finansowanego ze środków Unii Europejskiej w ramach Europejskiego Instrumentu Sąsiedztwa i Partnerstwa.

Niniejsza publikacja została stworzona przy pomocy Unii Europejskiej. Wyłączną odpowiedzialność za zawartość niniejszej publikacji ponosi Marcin Maciejewski oraz w żaden sposób nie może być ona postrzegana jako odzwierciedlenie poglądów Unii Europejskiej.

Literatura || Література

- [1] Branson A., Hauer T., McClatchey R., Rouglin D., Shamdasani J.: A Data Model for Integrating Heterogeneous Medical Data in the Health-e-Child Project CCS Research Centre Stud Health Technol Inform, 2008.
- [2] Ciosa K.J., Moore G.W.: Uniqueness of medical data mining, *Artificial Intelligence in Medicine* 26, 2002, 1–24.
- [3] Hariz M., Adnan M., Husain W.: Data Mining for Medical Systems: A Review, *Proceedings International Conference on Advances in Computer and Information Technology - ACIT 2012*.
- [4] Healey M.P., Jacobson J.E.: *Common Medical Diagnoses: An Algorithmic Approach*, Saunders, 2006.
- [5] Miller M., Rabczenko D., Wojtyński B.: *Epidemiologia w zdrowiu publicznym*, Wydawnictwo Lekarskie PZWL, 1/2010.
- [6] Mitchell T.M.: *Machine Learning* McGraw-Hill, 2011.
- [7] Morris F. C.: *Computer Medical Databases The First Six Decades (1950–2010)*, Springer, 2011.
- [8] Nguyen H.P., Kreinovich V.: Fuzzy Logic and its Applications in Medicine
- [9] Phuong N. H., Kreinovich V.: Fuzzy logic and its applications in medicine. *Int. J. Med. Inform.*, 2001.
- [10] Salski A., Holsten B., Trepel M.: A fuzzy approach to ecological modelling and data analysis, *Handbook of Ecological Modelling and Informatics*, Wit Press, 01/2009.
- [11] Shortliffe, E.H., Cimino J. J.: *Biomedical Informatics Computer Applications in Health Care and Biomedicine*, Springer, 2014.
- [12] Tarski A.: *Introduction to Logic and to the Methodology of Deductive Sciences*, Courier Dover Publications, Oxford University Press, 1994.
- [13] Wojtusik J.: *Semantic Data Types in Machine Learning from Healthcare Data*, *Proceedings of the International Conference on Machine Learning and Applications (ICMLA)*, Florida, 2012.
- [14] Zadeh L.A.: Fuzzy sets, *Information and Control*, 1965.
- [15] <http://guidetodatamining.com/>
- [16] <http://www.cdisc.org/>
- [17] <http://www.cdisc.org/system/files/all/article/image/jpeg/newdataset.jpg>
- [18] <http://www.hl7.org/>
- [19] <http://i.msdn.microsoft.com/dynimg/IC117351.gif>
- [20] <http://www.openehr.org/>
- [21] http://www.openehr.org/files/what_is_openehr/multi_level_modelling.png
- [22] <http://www.who.int/classifications/icd/en/>

Mgr inż. Marcin Maciejewski
e-mail: m.maciejewski@pollub.pl

Doktorant na Wydziale Elektrotechniki i Informatyki Politechniki Lubelskiej. Zajmuje się przetwarzaniem danych w medycynie, urządzeniami do telemedycyny, przetwarzaniem obrazu i systemami mikroprocesorowymi. Autor i współautor prac z dziedziny bioinformatyki, symulacji komputerowej w biochemii oraz przetwarzania obrazów. Opiekun Koła Naukowego Microchip.



MagiŃtr inż. Marcin MaŃcewŃski
e-mail: m.maciejewski@pollub.pl

Aspirant kaŃedry Elektrotechniki ta Informatyki LubliŃskiej Politechniki. ZajmaŃsya obróbkaŃ daŃich w medycynie, przyŃstrojami dla telemedycyny, obróbkaŃ zobraŃeŃ i mikroprocesorŃnymi systemami. Autor i spwaŃtor robŃt w gaŃluŃzi bioinformatyki, biochemii, a taŃoŃoŃ komputerŃnego modelowania ta obróbkaŃ zobraŃeŃ. OpikuŃ naukowŃego gurtka Microchip.