НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФАКУЛЬТЕТ ПРИКЛАДНОЇ МАТЕМАТИКИ

КАФЕДРА СИСТЕМНОГО ПРОГРАМУВАННЯ І
СПЕЦІАЛІЗОВАНИХ КОМП'ЮТЕРНИХ СИСТЕМ

«На правах рукопису»
УДК 681.3.06

«До захисту допущено»
Завідувач кафедри СПСКС

_____Віталій РОМАНКЕВИЧ
(підпис) (ім'я, прізвище)
"___"_____2020р.

Магістерська дисертація

на здобуття ступеня магістра

зі спеціальності 123 Комп'ютерна інженерія

на тему: Система багатофакторної аутентифікації користувачів комп'ютерних систем _____

Виконав: студент ІІ курсу, групи КВ-93мп
Тодорів Андрій Дмитрович _____ _____
(прізвище, ім'я, по батькові) (підпис)

Науковий керівник: професор д.т.н., проф. Ігор ТЕРЕЙКОВСЬКИЙ _____
_____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Рецензент_____ _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант з нормоконтролю доцент, с.н.с.,к.т.н. Юлія БОЯРІНОВА _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.
Студент _____
(підпис)

Київ – 2020 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Факультет прикладної математики

Кафедра системного програмування і спеціалізованих комп'ютерних систем

Рівень вищої освіти – другий (магістерський)
за освітньо-професійною програмою
Спеціальність 123 Комп'ютерна інженерія

ЗАТВЕРДЖУЮ
Завідувач кафедри СПСКС
_____ Віталій РОМАНКЕВИЧ
(підпис) (ініціали, прізвище)
«___»_____2020_р.

ЗАВДАННЯ
на магістерську дисертацію студенту
Тодорів Андрій Дмитрович
(прізвище, ім'я, по батькові)

1. Тема дисертації «Система багатофакторної аутентифікації комп'ютерних систем»,
науковий керівник дисертації с.н.с.,д.т.н. Терейковський Ігор Анатолійович,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені наказом по університету від «12» листопада 2020 р. №3298-С

2. Термін подання студентом дисертації 14 грудня 2020 р.

---

3. Об'єкт дослідження трансформація антропометричних показників в комп'ютерну форму

4. Предмет дослідження механізми розпізнавання образів.

5. Перелік завдань, які потрібно розробити аналіз існуючих методів розпізнавання образів; обгрунтування доцільності розробки алгоритму аутентифікації користувачів комп'ютерних систем на основі візуальних та голосових антропометричних показників; розробка алгоритмів голосової та візуальної ідентифікації користувачів комп'ютерних систем; аналіз розробленого алгоритму, його модифікація його порівняння з існуючими.

6. Перелік ілюстративного матеріалу - презентація

_____

_____

7. Перелік публікацій - 2 тез

_____

_____

8. Дата видачі завдання 5 листопада 2019 р. _____

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Затвердження теми	11.10.2019	
2	Збір та дослідження літератури	19.09.2020	
3	Аналіз існуючих рішень	23.09.2020	
4	Зміст та вступ	25.09.2020	
5	Розробка програмної моделі	05.10.2020	
6	Реферат	09.10.2020	
7	Перший розділ	09.10.2020	
8	Другий розділ	23.10.2020	
9	Третій розділ	04.11.2020	
10	Четвертий розділ	16.11.2020	
11	Редагування та перевірка роботи	20.11.2020	
12	Попередній розгляд магістерської дисертації на кафедрі	27.11.2020	

Студент _____ Андрій ТОДОРІВ

(підпис)

Науковий керівник дисертації _____ Ігор ТЕРЕЙКОВСЬКИЙ

РЕФЕРАТ

Актуальність теми

Вирішення проблеми захисту корпоративних даних в XXI столітті вийшло за рамки фізичної взаємодії з працівниками, у зв'язку з переходом шуканої інформації в комп'ютерний формат. Дана особливість сформувала потребу у розробці та імплементації нових механізмів захисту корпоративних даних.

Запропонована система аутентифікації користувачів комп'ютерних систем, розроблена на основі технологій нейронних мереж, надає можливість ідентифікації користувачів на основі індивідуальних антропометричних візуальних та голосових показників суб'єкта, з метою запобігання викраденню корпоративних даних, та ідентифікації злочинних суб'єктів.

Об'єктом дослідження є трансформація антропометричних показників в комп'ютерну форму.

Предметом дослідження є механізми розпізнавання образів.

Метою роботи є покращення можливостей методів біометричної ідентифікації суб'єктів шляхом розробки нової архітектури на базі нейронних мереж.

Методи дослідження. Порівняння існуючих алгоритмів за критеріями точності, швидкодії, ресурсних затрат, надійності, з метою імплементації та подальшої модифікації в системі корпоративного контролю.

Наукова новизна полягає у розробці нового механізму ідентифікації суб'єктів що поєднує у собі алгоритми голосової та візуальної ідентифікації суб'єктів.

Практична цінність полягає у можливості застосування даної системи в корпоративних умовах з метою запобігання витоку даних та ідентифікації злочинних суб'єктів. Низька ресурсозатратність сприяє застосуванню розробленого алгоритму в високонавантажених системах.

Структура та обсяг роботи. Магістерська дисертація складається з вступу, чотирьох розділів, висновків та додатків.

У вступі аналізується проблема захисту корпоративних даних. Обгрунтовується перспективність застосування механізмів біометричної голосової та візуальної ідентифікації суб'єктів для її вирішення. Досліджуються алгоритми біометричної ідентифікації.

У першому розділі описуються існуючі алгоритми розпізнавання візуальних та голосових образів.

У другому розділі досліджується доцільність застосування існуючих алгоритмів голосової та візуальної біометричної ідентифікації, аналізуються та порівнюються існуючі архітектури розпізнавання образів.

У третьому розділі наводиться процес розробки алгоритмів візуальної та голосової біометричної ідентифікації користувачів

У четвертому розділі наводяться характеристики розробленої КС, результати тестування, відбувається дослідження системи на різних наборах даних, та її модифікація з метою досягнення поставленої точності.

У висновках стисло наводяться результати досліджень та розробки.

Ключові слова: біометрична ідентифікація, антропометричні показники, нейронні мережі, розпізнавання образів, алгоритм, трансформація голосового сигналу, обробка зображень.

ABSTRACT

Topic relevance

The solution to the problem of corporate data protection in the XXI century has gone beyond the physical interaction with employees, due to the transition of the required information into a computer format. This feature has formed the need to develop and implement new mechanisms for corporate data protection.

The proposed system of authentication of computer system users, developed on the basis of neural network technologies, provides the possibility of user identification on the basis of individual anthropometric visual and voice indicators of the subject, in order to prevent theft of corporate data and identification of criminal entities.

The object of study is the transformation of anthropometric indicators into a computer form.

The subject of study is the mechanisms of pattern recognition.

The goal of this work is to improve the capabilities of biometric identification methods of subjects by developing a new architecture based on neural networks.

Study methods. Comparison of existing algorithms on the criteria of accuracy, speed, resource costs, reliability, in order to implement and further modify the corporate control system.

The scientific novelty is the development of a new mechanism for identifying subjects that combines algorithms for voice and visual identification of subjects.

The practical value lies in the possibility of using this system in a corporate environment in order to prevent data leakage and identification of

criminal entities. Low resource consumption contributes to the application of the developed algorithm in highly loaded systems.

Structure and scope of work. The master's dissertation consists of an introduction, four chapters, conclusions and appendices.

The introduction analyzes the problem of corporate data protection. The prospects of using the mechanisms of biometric voice and visual identification of subjects for its solution are substantiated. Biometric identification algorithms are investigated.

The first section describes the existing algorithms for recognizing visual and voice images.

The second section investigates the feasibility of using existing algorithms for voice and visual biometric identification, analyzes and compares existing image recognition architectures.

The third section describes the process of developing algorithms for visual and voice biometric user identification

The fourth section presents the characteristics of the developed COP, the test results, the system is studied on different data sets, and its modification in order to achieve the specified accuracy.

The conclusions summarize the results of research and development.

Key words: biometric identification, anthropometric indicators, neural networks, pattern recognition, algorithm, voice signal transformation, image processing.

ЗМІСТ

СПИСОК ТЕРМІНІВ, СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ

ПММ – приховані моделі Маркова (hidden Markov models)

НМ – нейронна мережа (Neural Network)

PCA – аналіз основних компонентів (Principal Component Analysis)

CNN – Convolution Neural Network

DNN – Deep Neural Network

AI – штучний інтелект (Artificial Intelligence)

WER – показник помилкового розпізнавання (Word Error Rate)

ВСТУП

Проблема біометричної ідентифікації суб'єктів постала перед людством ще в другій половині XX століття, спершу в якості задач розпізнавання, а згодом, в рамках задач ідентифікації суб'єктів.

Розробки в сфері біометричної ідентифікації застосовуються в широкому колі галузей:

- Військова;
- Корпоративна;
- Урядова
- Наукова

Варто виділити найбільш знакові розробки в даних сферах, зокрема:

- військово-урядова китайська система біометричної візуальної ідентифікації, що імплементована в провінції Сінцьзянь, задля забезпечення класової сегрегації мусульманського населення етнічних меншин уйгурів.
- Face ID – добре відома технологія візуальної біометричної ідентифікації, застосована в продуктах фірми Apple задля впровадження системи обмеженого доступу до пристроїв.

Технічне рішення що представлено в даній роботі виходить за рамки відомих розробок і пропонує об'єднання різних технологій біометричної ідентифікації в кінцевий програмний продукт, що здійснюватиме процес ідентифікації на основі кількох біометричних показників суб'єкта.

Запропонована розробка має перспективи застосування в галузі корпоративної безпеки. Зокрема, незалежні статистичні дані 2019 року доводять: 92,5% світових компаній стикалися з витоком даних:

- 55,2% випадків - причина витоку - цілеспрямовані дії співробітників;

- 23,6% випадків – необережне поводження з інформацією;
- 21% випадків – дії хакерів та вірусів.

Згідно цієї ж статистики:

- 37% працівників компаній не проти продажу корпоративної інформації фірмам конкурентам;
- 26% працівників розвивають свій власний бізнес за рахунок корпоративних ресурсів.

В свою чергу:

- 59% респондентів вважають найбільш пріоритетною загрозою внутрішню - працівників, що можуть легко скористатися корпоративною інформацією в корисливих цілях;
- 33% сприймають необережність працівників при користуванні даними за основний фактор;
- 8% вбачають провину хакерів у злитті даних.

Три найважливіші причини, завдяки яким співробітник може скоїти крадіжку:

- Відсутність систем контролю персоналу і переміщення інформації - 41,6%;
- Відсутність персональної відповідальності - 34,7%;
- Велике коло осіб має доступ до інформації - 23,7%.

У 46% випадків інциденти ведуть до безповоротної втрати даних. Причини:

- Цілеспрямоване шкідництво - 36,3%;
- Необдумані дії персоналу - 31,9%;
- Дії третіх осіб ззовні - 31,3%.

Згідно статистиці, людський фактор – є ключовим у питанні витоку інформації, контроль цього явища потребує нових технічних засобів задля нівелювання інформаційних та фінансових втрат в індустрії.

Імплементація нових алгоритмів контролю над працівниками призведе до формування загальнодоступної бази даних ненадійних працівників, персон нон грата, що раніше спостерігалися за крадіжками корпоративної інформації, з метою подальшого їх не наймання у структури.

Одним з таких алгоритмів є запропоноване у даній роботі технічне рішення багатофакторної системи аутентифікації користувачів на основі антрометричних показників.

Статистично обгрунтовано, що використання такого рішення вирішить проблему відсутності систем контролю над персоналом і призведе до зменшення показників витоків інформації у 41,6% випадків, а також створить можливість ідентифікації ненадійних працівників, які є причиною 55,2% інформаційних крадіжок корпоративної інформації.

1. АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ ТА ОБГРУНТУВАННЯ ТЕМИ МАГІСТЕРСЬКОЇ ДИСЕРТАЦІЇ

1.1. Загальний опис проблеми аутентифікації користувачів на основі антропометричних даних

В даній роботі проблему біометричної аутентифікації користувача можна умовно розділити на дві підзадачі: задачу візуальної аутентифікації на основі індивідуальних показників обличчя, та задачу аутентифікації на основі голосового сигналу користувача.

Задача візуальної ідентифікації на основі зображення обличчя була вперше сформульована професором Техаського університету Остіна Вуді Бледсо наприкінці 1950-х років. В розумінні Бледсо, задача ідентифікації зводилася до розпізнавання десяти обличь, та вже 1965 року розробками вченого зацікавилося ЦРУ США , а також, зрозумілими стали можливі наслідки впровадження такої розробки – соціально-расова сегрегація (на прикладі ізоляції уйгурів на основі расових антропометричних показників).

Задача голосової ідентифікації була поставлена ще в 60-х роках минулого століття, ще в рамках задачі розпізнавання голосу. При цьому, перші відомі методи вирішення стали наявні лише на початку 90-х років, через розвиток апаратних технологій, що надавали можливості обробки великих наборів даних. Розробки, спрямовані на розпізнавання голосового сигналу поширені у всіх сучасних інформаційних галузях, починаючи від методів голосового набору тексту у телефоні, і закінчуючи військовою галуззю, де дана технологія використовується задля аналізу сенсу телефонних розмов, з метою наступної ідентифікації деструктивного контенту та попередження терористичних актів, тощо. Можливість подальшої ідентифікації терористичних суб’єктів на основі

антропометричних ознак їх голосу є перспективним методом та прикладом застосування запропонованої технології.

Вдосконалення можливостей моделей розпізнавання, їх комбінація в комп'ютерних системах нового гатунку, є комплексною проблемою, сукупна користь розв'язання якої призведе до появи нових засобів корпоративного контролю за персоналом, а також стане будівельним блоком для методів захисту конфіденціальної інформації в умовах підприємства.

1.2.Особливості візуальної біометричної ідентифікації

Візуальна ідентифікація - це техніка розпізнавання, яка використовується для виявлення обличь осіб, чиї зображення зберігаються у наборі даних. Незважаючи на те, що інші методи ідентифікації можуть бути більш точними, розпізнавання обличь завжди залишалося важливим напрямком досліджень через його невтручальний характер і тому, що це легкий метод особистої ідентифікації.

Перші напрацювання в даній галузі були сформовані на початку 60-х років минулого століття, спочатку в якості задачі розпізнавання зображених символів.

Вуді Бледсо і його колега Ібен Браунінг (Iben Browning) - винахідник-ерудит, авіаінженер і біофізик - придумали метод, який згодом став відомий як метод кортежів (n-tuple).

Вчені почали з проектування надрукованого символу - скажімо, букви Q - на прямокутну сітку з клітин на зразок розлініяного аркуша паперу. Кожній клітинці-осередку присвоювався двійковий номер в залежності від наявності або відсутності в ній частині символу: 0 - для порожньої клітки, 1 - для заповненою. Потім осередки випадковим чином групувалися у впорядковані пари, як набори координат. Теоретично групи могли включати будь-яку кількість осередків, звідси і назва методу. Далі

за допомогою декількох математичних дій система присвоювала сітці символу унікальне значення. А при зіткненні з новим символом сітка цього символу порівнювалася з іншими в базі даних до тих пір, поки не знаходився найближчий збіг.

Суть методу полягала в тому, що він дозволяв розпізнавати безліч варіантів одного і того ж знака: більшість Q, як правило, отримували досить схожі результати в порівнянні з іншими Q. Найкраще процес працював з будь-яким шаблоном, а не тільки з текстом. За словами Роберта С. Бойера (Robert S. Boyer), математика, метод кортежів допоміг визначити область розпізнавання шаблонів. Це був один з перших кроків до питання: «Як запрограмувати машину робити те, що роблять люди?".

В 1963 році задача ідентифікації десяти людей була неймовірно амбітною.

За її реалізацію взялася компанія Panoramic, головним інженером якої був Бледсо. Стрибок від розпізнавання написаних символів до розпізнавання осіб був гігантським. Хоча б тому що не було ні стандартного методу оцифровки фотографій, ні існуючої бази цифрових зображень, на яку можна було б спиратися. Сучасні дослідники можуть навчати свої алгоритми на мільйонах селфі у вільному доступі, а Panoramic довелося б будувати базу даних з нуля. Була серйозніша проблема: тривимірні особи людей на відміну від двомірних знаків не статичні. Зображення одної людини можуть відрізнятися за поворотом голови, інтенсивністю освітлення і ракурсом, а також в залежності від віку, зачіски і настрою, виразу обличчя - на одній фотографії людина може виглядати безтурботною, на інший - стурбованою.

Задачею команди було скорегувати варіативність і впорядкувати зображення, які вони порівнювали.

Однією з основних машин була CDC 1604 зі 192 КБ оперативної пам'яті - приблизно в 21 000 разів менше, ніж у звичайного сучасного смартфона.

З самого початку Бледсо повністю усвідомлював ці складності, тому розбив дослідження на частини і доручив їх різним співробітникам.

Робота над оцифруванням зображень проходила наступним чином. Дослідник знімав чорно-білі фотографії учасників проекту на 16-міліметрову плівку. Потім використовував скануючий пристрій, який розробив Браунінг, щоб перетворити кожен знімок в десятки тисяч точок даних. Кожна точка повинна була мати значення інтенсивності світла в діапазоні від 0 (найтемніша) до 3 (найсвітліша) - в певному місці на знімку. Виходило дуже багато точок для одноразової обробки комп'ютером, тому дослідник написав програму NUBLOB, яка нарізала зображення на зразки випадкового розміру і обчислювала для кожного унікальне значення - на зразок тих, що присвоювалися за методом кортежів.

Над нахилом голови працювали Вуді, Хелен Чан Вульф і ще один дослідник. Спочатку вчені намалювали серію пронумерованих маленьких хрестиків на лівій стороні обличчя учасника експерименту - від вершини чола до підборіддя. Потім зробили два портрета, на одному з яких людина дивилася вперед, а на іншому – була повернута на 45 градусів. Проаналізувавши розташування хрестиків на цих двох зображеннях, екстраполювали дані на знімок особи з поворотом на 15 або 30 градусів. Завантажували в комп'ютер чорно-білу картинку розміченої особи, а на виході отримували дивовижно точну, точкову модель-портрет, що обертається.

Рішення дослідників були оригінальними, але недостатньо ефективними. Через тринадцять місяців після початку роботи команда Panoramic визнала, що їм не вдалося навчити машину розпізнати хоча б одну особу, не те що десять.

Ріст волосся, виразу обличчя і ознаки старіння - цей потрійний виклик був «колосальним джерелом мінливості», - написав Вуді в березні 1964 року в звіті про виконану роботу для King-Hurley. Поставлена задача «виходить за рамки поточного стану галузі розпізнавання образів і сучасних комп'ютерних технологій». При цьому Вуді рекомендував фінансувати більше досліджень, щоб спробувати знайти «абсолютно новий підхід» до вирішення проблеми розпізнавання осіб.

Протягом наступного року Вуді прийшов до висновку, що найбільш багатообіцяючий підхід до автоматизованого розпізнавання осіб - той, який звужує область до взаємозв'язків між головними елементами: очима, вухами, носом, бровами, губами.

Система, яку він запропонував, була схожа на метод французького кримінолога Альфонса Бертильона, який він створив в 1879 році. Бертільон описував людей на основі 11 фізичних вимірювань, включаючи довжину лівої ноги і довжину від ліктя до кінця середнього пальця. Ідея полягала в тому, що якщо провести досить вимірювань, то кожна людина стане унікальною. Метод був трудомістким, але працював: за допомогою нього в 1897 році, задовго до широкого поширення дактилоскопії, французькі жандарми ідентифікували серійного вбивцю Жозефа Ваше.

Протягом 1965 року Panoramic намагалася створити повністю автоматизовану систему Бертильона для ідентифікації осіб. Команда намагалася розробити програму, яка могла б визначати носи, губи та інше за допомогою світлих і темних ділянок на фотографії. Але їх спіткала невдача.

Тоді Вуді і Вульф зайнялися вивченням того, що вони назвали «людино-машинним» підходом до розпізнавання осіб - методом, який включив в рівняння трохи людського участі.

До проекту Вуді привернув свого сина Грегорі і його друга - їм дали 122 фотографії, на яких було зображено близько 50 осіб. Хлопці зробили

22 вимірювання кожної особи, включаючи довжину вуха і ширину рота. Потім Вульф написала програму для обробки даних.

Машина навчилася зіставляти кожен комплект вимірювань з фотографією. Результати були скромними, але незаперечними: Вульф і Вуді довели, що система Бертильона теоретично працездатна.

Їх наступним кроком, в кінці 1965 року народження, було створення більш масштабної версії того ж експерименту, щоб зробити «людину» більш ефективною в їх системі «людина-машина». На гроші King-Hurley вони придбали планшет RAND - пристрій вартістю 18 000 доларів, який виглядав як планшетний сканер зображень, а працював як iPad. За допомогою стилуса дослідник малював на планшеті і на виході отримував комп'ютерне зображення досить високої роздільної здатності.

Через планшет RAND провели нову партію фотографій, підкреслюючи стилусом ключові елементи особи. Цей процес хоча і був складним, але проходив набагато швидше, ніж раніше: дані ввели приблизно для 2 000 знімків, включаючи як мінімум два зображення кожної особи. В годину обробляли близько 40 знімків.

Навіть при такій, більшій вибірці команда Вуді насилу долала звичайні перешкоди.

Як і раніше не була вирішена проблема з посмішками, які «спотворюють обличчя і кардинально змінюють міжлицеве вимірювання», а також зі старінням.

Рисунок 1.1 – основні біометричні характеристики обличчя

При спробі зіставити фотографію Вуді 1945 року з фотографією 1965 року (рис. 1.1) система збивалася з пантелику. Вона не бачила великого подібності між молодою людиною з широкою посмішкою і густими темними волоссям і людиною більш старшого віку з похмурим виразом обличчя і поріділої шевелюрою.

У 1967 році Вуді взявся за останнє завдання, пов'язане з розпізнаванням патернів особи. Метою експерименту було допомогти правоохоронним органам швидко аналізувати бази даних заарештованих в пошуках збігів.

Як і раніше, фінансування проекту, судячи з усього, надійшло від уряду США. У документі 1967 року, що розсекреченому ЦРУ в 2005 році, згадується «зовнішній контракт» на систему розпізнавання осіб, що дозволило б в сто разів скоротити час пошуку.

Основним партнером Вуді по проекту був Пітер Харт (Peter Hart), інженер-дослідник Лабораторії прикладної фізики Стенфордського науково-дослідного інституту. (Зараз відомий як SRI International. Інститут відокремився від Стенфордського університету в 1970 році через розбіжності в кампусі з приводу сильної залежності інституту від військового фінансування.)

Вуді і Харт почали з бази даних з близько 800 знімків - по два знімка «400 дорослих чоловіків європеоїдної раси». Сфотографовані розрізнялися за віком і поворотом голови. За допомогою планшета RAND вчені зафіксували 46 координат для кожної фотографії, в тому числі п'ять значень для кожного вуха, сім - для носа і чотири - для кожної брови. На базі попереднього досвіду Вуді по нормалізації варіацій зображень застосували математичне рівняння, щоб «повернути» голови в анфас. Потім, для обліку різниці в масштабах, збільшили або зменшили кожне зображення до стандартного розміру, де опорною метрикою була відстань між зіницями.

Завдання системи полягало в тому, щоб запам'ятати одну версію кожної особи і використовувати її для ідентифікації іншої.

Вуді і Харт запропонували машині один з двох коротких шляхів. При першому, відомому як груповий збіг, система розділяла обличчя на риси - ліва брова, праве вухо і так далі - і порівнювала відносні відстані між ними (рис. 1.2).

Другий підхід грунтувався на байєсівській теорії прийняття рішень, де машина використовувала 22 вимірювання, щоб зробити загальне обґрунтоване припущення.

За підсумками обидві програми впоралися із завданням приблизно однаково добре. А також виявилися кращими за суперників-людей. Коли Вуді і Харт попросили трьох осіб зіставити підгрупи з 100 осіб, навіть найшвидшому з них знадобилося шість годин. Комп'ютер CDC 3800

виконав аналогічне завдання приблизно за три хвилини, домігшись стократного скорочення часу. Люди краще справлялися з поворотами голови і поганою якістю фотозйомки, але комп'ютер «значно перевершував» в плані визначення вікових змін.

Дослідники прийшли до висновку, що машина «домінує» або «майже домінує» над людиною. Це був найбільший успіх Вуді в його дослідженнях по розпізнаванню обличчя. Це була також його остання робота по цій темі, яка ніколи не була опублікована «в інтересах держави», про що Вуді і Харт дуже шкодували.

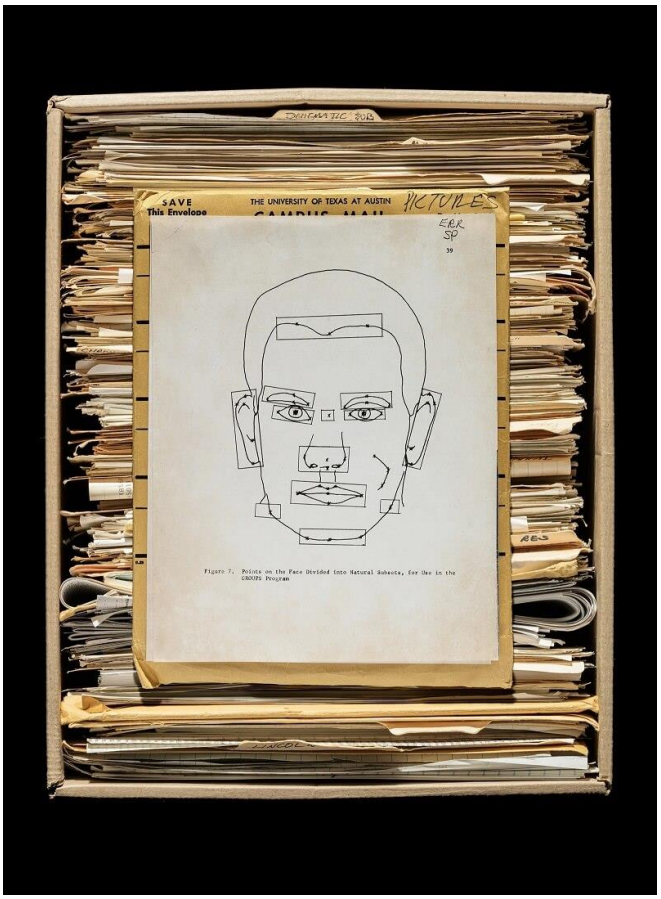


Рисунок 1.2 – поділ обличчя на риси

У 1973 році японський вчений-програміст Такео Канаді (Takeo Kanade) зробив великий стрибок в технології розпізнавання осіб.

На основі бази даних з 850 оцифрованих фотографій зі Світовою організацією виставки в Суіте (Японія) в 1970 році Канаді розробив програму, яка могла отримувати риси обличчя - ніс, рот і очі - без участі

людини. Канаді вдалося здійснити мрію Вуді про виключення людини з системи «людина-машина».

«Тільки за останні 10 років технологія розпізнавання осіб навчилася працювати з недосконалостями», - говорить Аніл К. Джейн (Anil K. Jain), вчений-програміст Мічиганського державного університету і співредактор Керівництва по розпізнаванню осіб (Handbook of Face Recognition).

Майже всі проблеми, з якими стикався Бледсо, відпали. Сьогодні наявний невичерпний запас оцифрованих зображень. «Через соціальні мережі ви можете отримувати стільки знімків особи, скільки захочете», - говорить Джейн. А завдяки досягненням в області машинного навчання, обсягом пам'яті і обчислювальної потужності комп'ютери ефективно самонавчаються. З огляду на кілька простих правил, вони можуть аналізувати величезні обсяги даних і створювати шаблони практично для чого завгодно, починаючи від людського обличчя і закінчуючи пакетом чіпсів - ніяких вимірів за допомогою планшета RAND або методу Бертильона більше не потрібно.

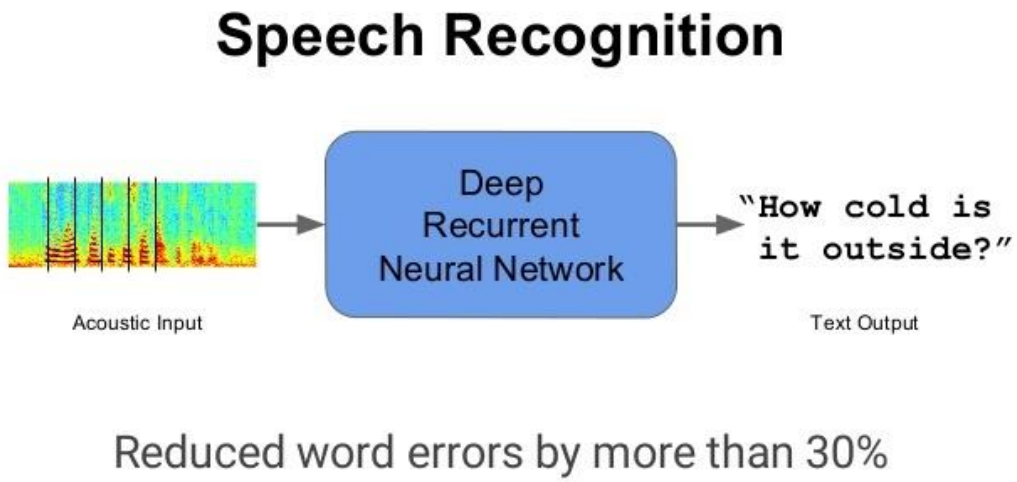
1.3.Особливості голосової біометричної ідентифікації

Перші напрацювання в галузі голосової ідентифікації були створені у вигляді вирішення задачі розпізнавання мовлення. Вже наприкінці 80-х років XX століття створені перші системи розпізнавання на базі HMM. Розпізнавання мовлення є бажаною технологією, сфери застосування якої поширюються від цивільного та громадського, до приватного та військового інтересу. Поки алгоритми штучного інтелекту збирають та аналізують наші побажання, які матеріалізують на пристроях у вигляді контекстної реклами, військові бази в пасивному режимі збирають та класифікують інформацію про потенційних вбивць, терористів, торговців зброєю та наркотиками. Дана практика забезпечує запобігання злочинній активності з мінімальними витратами часу та ресурсів.

Можливість побудови моделей алгоритмів розпізнавання з’явилася відносно недавно, через апаратні обмеження, що були подолані появою нових архітектур на зламі XX-XXI ст.

Вхідними даними виступають голосові команди, представлені у вигляді акустичного аудіо сигналу, що в подальшому має бути трансформований в цифрову форму для забезпечення можливості аналізу з боку НМ.

Таким чином, проблема класифікується за кількість окремих наборів індивідуальних формант. Їх кількість визначає набір класів, що будуть задіяні у вирішенні проблеми. Задача розпізнавання за класами постала 50 років тому на прикладі системи розпізнавання мовлення “Watermelon”, що була здатна до розпізнавання одного слова – watermelon. Дана можливість була надана за рахунок спеціального розміщення голосних у слові, програма мала лише один клас і два випадки, що дозволяло побудувати рішення на повністю аналоговій базі. З цього випливає, що збільшення кількості класів розпізнавання веде до потреби збільшення апаратних потужностей задіяних в процесі аналізу, а також призводить до



збільшення аналітичного часу.

Рисунок 1.3 – розпізнавання голосового сигналу РНМ

Наприкінці 90-х років XX століття, існуючі алгоритми та моделі надавали можливість класифікації до 100 слів, що було недостатнім

показником для безперешкодного розпізнавання чистого людського мовлення. Задача була непридатною для розв'язання з кількістю класів більше тисячі, і в умовах, коли слова були сказані без пауз. Кожна нова умова – наявність шуму, зміна швидкості мовлення, емоційного забарвлення голосу породжувала створення нових класів розпізнавання і унеможливлювала аналіз в даних апаратних умовах.

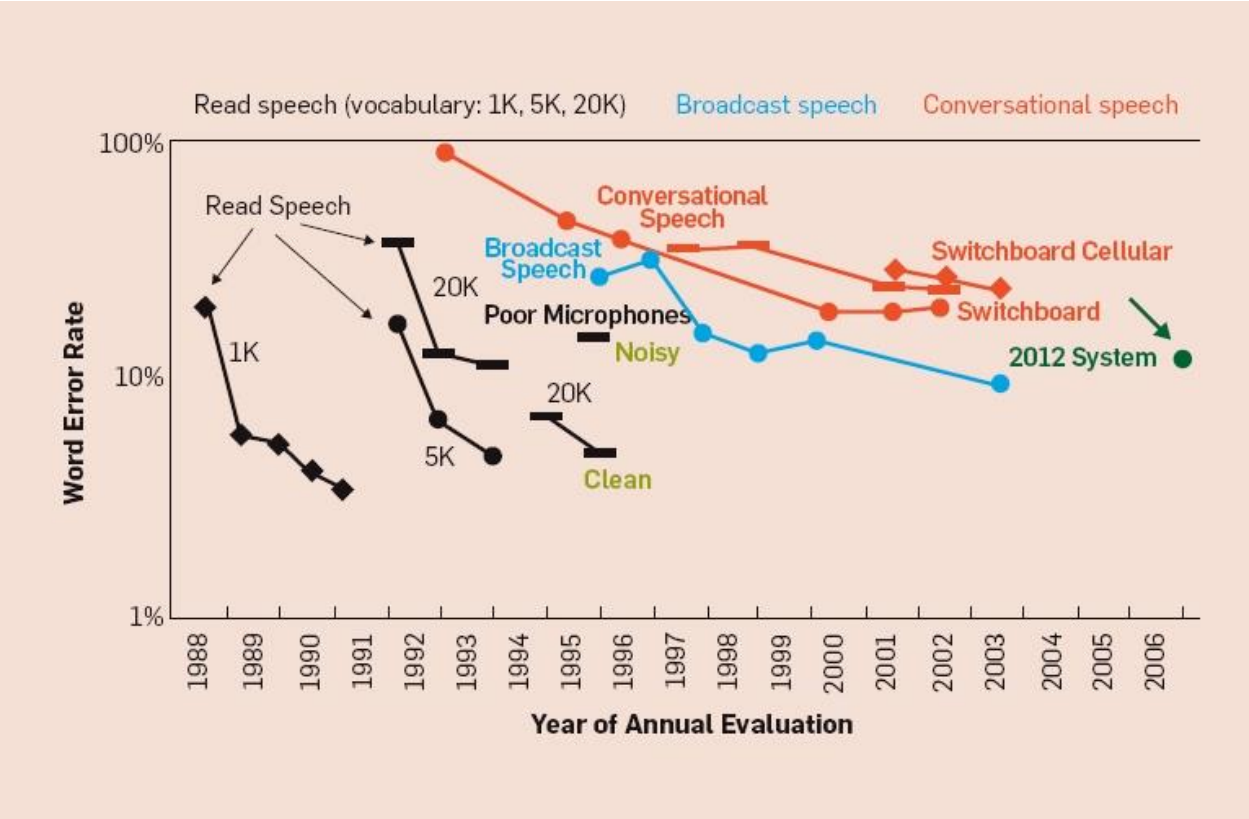


Рисунок 1.4 – розвиток якості систем розпізнавання

Головною метрикою в питанні розпізнавання мовлення є WER. WER – показник помилкового розпізнавання, відсоток неправильно розпізнаних слів. Згідно рис. 1.4, 1.5, окремою проблемою постає розпізнавання так званого «розмовного мовлення», яке на відміну від «мовлення читання» демонструє гірші показники WER. Даний фактор створює проблеми для військових досліджень, людське «розмовне мовлення», забарвлене шумом, при трансформації в цифрову форму на стільниковому телефоні проходить крізь смуги вузької пропускної здатності, що спотворює вхідні дані, і як наслідок, унеможливлює подальший аналіз.

Показники що впливають на WER:

- Розмір словника
- Якість запису
- Наявність звукових завад
- Наявність пауз між словами
- Особливості конкретної мови

Протягом 1999-2010 років якість розпізнавання залишалася сталою.

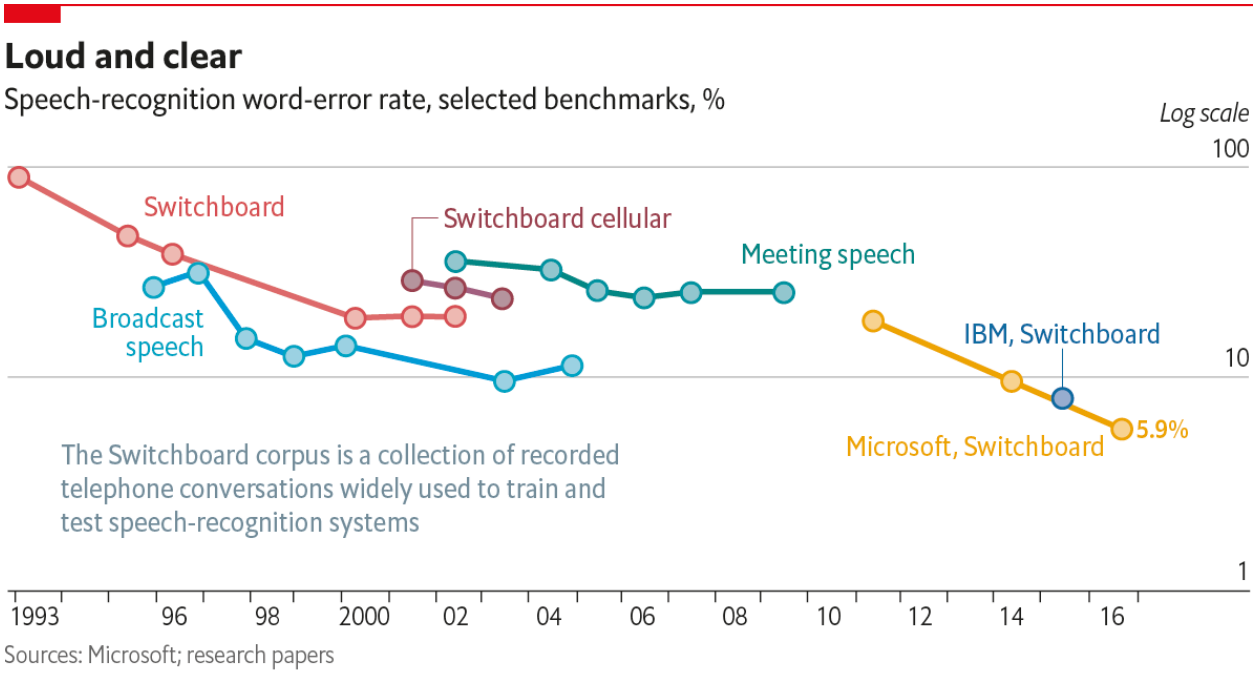


Рисунок 1.5 – розвиток якості систем розпізнавання «чистого» мовлення

Нові апаратні засоби надали можливість швидшого розпізнавання, та забезпечили системи підтримкою великих словників. Але якість розпізнавання залишалася сталою. Нові генерації алгоритмів також вимагали часових затрат на тренування HM, заради задоволення умовам WER. Рис. 1.6 зображує недоліки старих технологій в порівнянні з DNN.

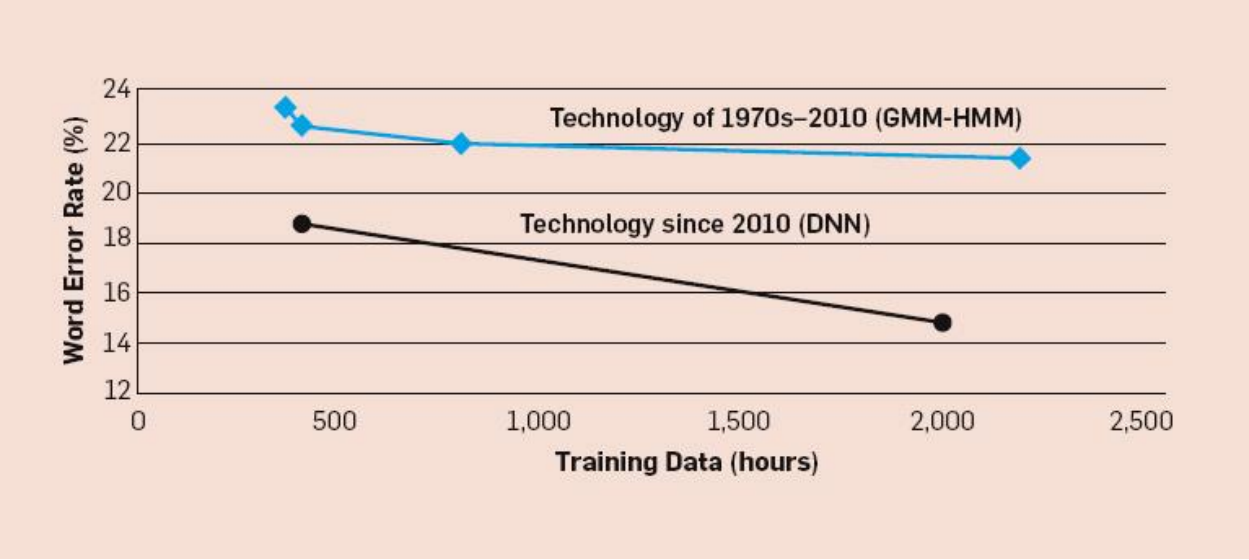


Рисунок 1.6 – показники WER для різних типів технологій

Так як розпізнавання мовлення не є ключовою задачею роботи, аналіз голосу можна звести до дослідження індивідуальних голосових показників людського індивіда – голосових формант. Форманти – це тембральні показники голосового забарвлення ,що є індивідуальними для кожної людини.

1.4. Висновки

Зважаючи на проведений аналіз, були сформовані вимоги до кінцевого програмного продукту – комп’ютерної системи біометричної ідентифікації користувача.

Завдання полягає у побудові двох підсистем біометричної ідентифікації:

- Система біометричної ідентифікації на основі голосового сигналу;
- Система біометричної візуальної ідентифікації.

Задля комбінації їх результатів з метою однозначного розпізнавання суб’єкта і імплементації розроблених рішень в кінцевий програмний продукт.

2. ВПРОВАДЖЕННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ ВИРІШЕННЯ ЗАДАЧ БІОМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ

2.1. Аналіз потенційних нейромережевих архітектур

Зважаючи на сучасний стан нейромережевих технологій, сформульовані наступні характеристики та умови доцільності застосування конкретних архітектур:

- Параметри вхідних даних;
- Вимоги до вихідної інформації;

- Апаратні обмеження;
- Обмеження часу навчання;
- Технічні обмеження нейронної мережі;
- Визначення сфери застосування.

Основні вимоги до освітніх даних:

- Достатня кількість параметрів, що описуються датасетом;
- Кількість прикладів датасету;
- Інформативність датасету;
- Необхідність підготовки вхідних даних;
- Повнота набору даних.

Загальні обмеження процесу навчання передбачені:

- Обмеження часу навчання НМ;
- Тип навчання - з викладачем або без;
- Потреба в автоматизації навчального процесу, що визначає загальну кількість отриманих параметрів навчання;
- Впровадження додаткових можливостей навчання під час або після процесу використання.
- Вимоги до якості навчання НМ, виражені у вибраних показниках - необхідних рівні помилкового розпізнавання.

На практиці вимоги до апаратного забезпечення визначаються максимальною кількістю прикладів, які мережа аналізує для досягнення необхідного рівня точності прийняття рішень. У свою чергу, точність характеризується значення максимальної та середньої похибки мережі на реальних даних, які загалом можуть виходити за рамки обсягу навчальних даних. Відповідно, завдання полягає в екстраполяції результатів навчання нейронної мережі за межі навчального набору. Слід зазначити, що обчислювальна потужність розробленої моделі залежить від її типу та алгоритму навчання.

Сформовані наступні обмеження щодо технічної реалізації НМ:

- швидкість прийняття рішень;
- інтеграція в існуюче обладнання;
- обсяг та складність реалізація програми.

Область застосування визначає шляхи застосування НМ. Насьогодні використання НМ для розпізнавання та оптимізації зображень набуло великого поширення. Слід зазначити, що системи розпізнавання зображень в принципі відрізняються від систем аналізу тексту тим, що мають принципово обмежену кількість виходів і комбінацій вхідних параметрів. У системах
аналізу тексту, ця кількість принципово необмежена.

Пристосованість мережі до автономної роботи визначає сферу її застосування. Для цього в архітектурі НМ необхідно передбачити здатність повністю автоматизувати навчальний процес. Спираючись на матеріали даної роботи можна стверджувати, що ключові задачі застосування НМ у галузі обслуговування програмного забезпечення техніко-економічних систем:

- розпізнавання образів;
- визначення оптимальних управлінських рішень;
- створення асоціативної пам'яті.

Перший напрямок включає завдання класифікації, кластеризації зображень та апроксимації функцій.

Другий напрямок включає завдання оптимального управління та задачі керування еталонною моделлю.

До третього напрямку належать завдання створення комп’ютерних інформаційних систем з пам’яттю, які розглядаються в даному проекті.

Крім того, галузь застосування технології залежить від її актуальності в конкретній сфері та спроможності розробленої мережі справлятися з конкретними задачами.

Доцільність НМ архітектур в конкретних задачах Таблиця 2.1

Умова	БШП	РБФ	SOM	АРТ	СНМ	PNN/ GRNN	Асоціа-тивні
Навчальні дані							
Допустимість шуму	1	0	1	-1	1	0	-1
Допустимість кореляції	1	1	1	1	1	1	-1
Повнота виборки	-1	1	1	-1	-1	1	0
Пропорційність прикладів	1	-1	-1	-1	-1	-1	0
Загальні обмеження процесу навчання							
Короткий термін навчання	-1	0	1	1	0	1	1
Представлення в навчальних прикладах очікуваного виходу	1	1	-1	-1	-1	1	1

Умова	БШП	РБФ	SOM	АРТ	СНМ	PNN/ GRNN	Асоціа-тивні
Автоматизація навчання	1	-1	0	1	1	1	0
Можливість донавчання	0	1	1	1	1	1	0
Якість навчання	1	0	0	1	1	1	1
Обчислювальні потужності							
Обсяг пам'яті	1	-1	-1	-1		-1	0
Екстраполяції результатів навчання	1	-1	-1	-1		-1	1
Незмінність результатів	1	1	0	1	1	1	0
Вихідна інформація							
Інтерпретації виходу у вигляді ймовірності	0	0	-1	-1	-1	1	0
Інтерпретації виходу у графічному вигляді	-1	-1	1	-1	-1	-1	-1
Можливість вербалізації	1	0	-1	-1	-1	0	-1
Обмеження технічної реалізації НМ							
Швидкості прийняття рішення	1	1	1	1	0	1	-1

Умова	БШП	РБФ	SOM	АРТ	СНМ	PNN/ GRNN	Асоціа-тивні
Обсяг програмної реалізації	-1	1	-1	0	-1	-1	0
Сфера застосування							
Системи розпізнавання образів	1	1	1	1	0	1	1
Системи аналізу тексту	-1	-1	1	0	1	0	-1
Системи управління	-1	-1	1	-1	-1	-1	1
Автономність функціонування	-1	-1	-1	1	1	-1	-1

Таблиця 2.1. відображає особливості та недоліки окремих систем НМ. Дані характеристики визначають доцільність застосування конкретних архітектур при вирішенні різних типів задач. Класифікація відбувається за наступною градацією:

- -1 – відсутність переваги при використанні системи в заданих умовах;
- 0 – звичайна продуктивність;
- 1 – наявність переваги.

Згідно таблиці 2.1 задачі розпізнавання візуальних образів обрано вирішувати за допомогою НМ глибокого навчання, а задачі розпізнавання голосу за допомогою рекурентних НМ

2.2 Визначення перспективних методів візуальної біометричної ідентифікації

Завдання розпізнавання обличчя - частина практичного застосування теорії розпізнавання образів. Вона складається з двох підзадач: ідентифікації та класифікації. Ідентифікація особистості активно використовується в сучасних сервісах, таких як Facebook, iPhoto. Розпізнавання обличчя використовується всюди, починаючи від FaceID в iPhone X, закінчуючи використанням при наведенні цілей у військовій техніці.

Людина розпізнає обличчя інших людей завдяки зоні мозку на кордоні потиличної і скроневої часток - веретеноподібної звивини. Ми розпізнаємо різних людей з 4-х місяців. Ключові особливості, які виділяє мозок для ідентифікації, - очі, ніс, рот і брови. Також людський мозок відновлює обличчя цілком навіть по половині і може визначити людину лише по частині особи. Все побачені особи мозок усереднює, а потім знаходить відмінності від цього усередненого варіанта. Тому людям європеоїдної раси здається, що всі, хто належить монголоїдної раси, на одну особу. А монголоїдам важко розрізняти європейців. Внутрішнє розпізнавання налаштоване на спектральному діапазоні осіб в голові, тому, якщо якоїсь частини спектра не вистачає даних, особа вважається за одне і теж.

Завдання по розпізнаванню обличчя вирішують вже більше 40 років. У них входить:

- Пошук і розпізнавання декількох осіб в потоці.
- Стійкість до змін особи, зачіски, бороди, окулярів, віку і повороту особи.
- Масштабованість даних для ідентифікації людини.

- Робота в реальному часі.

Системи розпізнавання обличь використовують комп’ютерні алгоритми, щоб виділити конкретні, відмінні деталі щодо обличчя людини. Ці деталі, такі як відстань між очима або форма підборіддя, потім перетворюються в математичне зображення і порівнюються з даними інших обличь, зібраними в базі даних. Дані про конкретне обличчя часто називають шаблоном обличчя, вони відрізняються від фотографії, оскільки розроблені таким чином, щоб включати лише певні деталі, які можна використовувати для розрізнення одного обличчя від іншого. Деякі системи розпізнавання обличь замість позитивної ідентифікації невідомої особи призначені для обчислення оцінки збігу ймовірностей між невідомою особою та конкретними шаблонами обличь, що зберігаються в базі даних. Ці системи запропонують кілька потенційних збігів, класифікованих за порядком вірогідності правильної ідентифікації, замість того, щоб просто повернути один результат. Системи розпізнавання обличчя відрізняються за здатністю ідентифікувати людей за таких складних умов, як погане освітлення, низька якість зображення та неоптимальний кут огляду.

Програмне забезпечення для розпізнавання обличчя засноване на здатності спочатку ідентифікувати обличчя, що саме по собі є технологічним рішенням. Подивившись у дзеркало, можна побачити, що на вашому обличчі є певні помітні орієнтири. Це вершини та долини, що складають різні риси обличчя. На обличчі людини є близько 80 вузлових точок.

Ось декілька вузлових точок, які вимірюються програмним забезпеченням.

- Відстань між очима;
- Ширина носа;
- Глибина очної ямки;
- Скули;

- Лінія щелепи;
- Підборіддя.

Ці вузлові точки вимірюються, щоб створити числовий код, рядок чисел, що представляє обличчя у базі даних. Цей код називається відбитком обличчя. Для завершення процесу розпізнавання програмному забезпеченню потрібно від 14 до 22 вузлових точок.

Різні вузлові точки (так звані орієнтири) існують на кожному обличчі - верхівці підборіддя, зовнішньому краю кожного ока, внутрішньому краю кожної брови тощо, як показано на малюнку нижче. Алгоритм машинного навчання здатний знаходити такі вузлові точки на будь-якому обличчі (рис. 2.1):

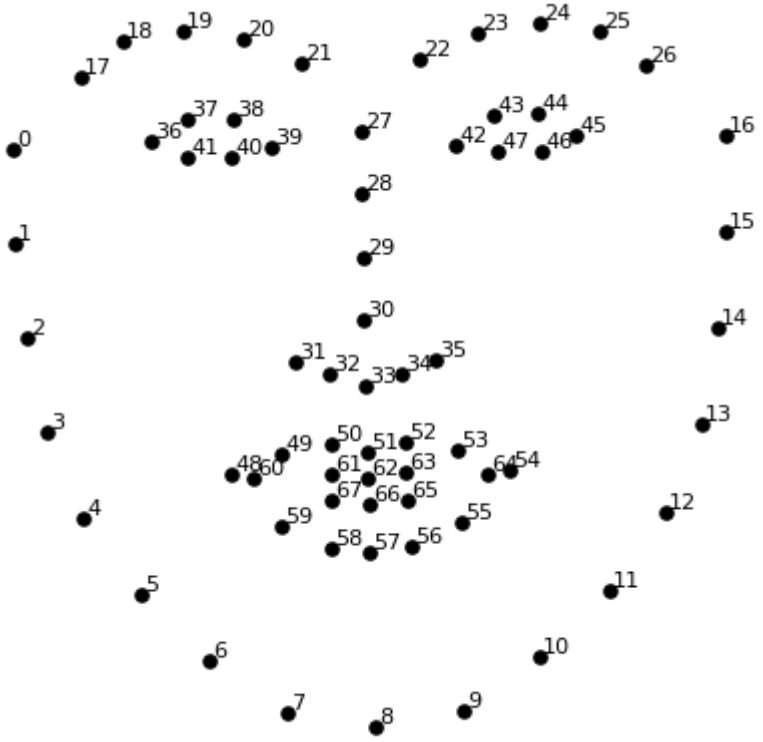


Рисунок 2.1 – вузлові лінії обличчя

Системи розпізнавання обличчя генерують так званий відбиток обличчя - унікальний код, що застосовується до однієї людини, - вимірюючи відстань між точками, як-от ширина носа людини.

Ці так звані «вузлові точки» - понад 80 точок, які перевіряє система розпізнавання обличчя, - об’єднуються математично для створення

відбитка обличчя, який потім може бути використаний для пошуку в базі даних ідентифікаційних даних. Кожен алгоритм машинного навчання приймає набір вхідних даних та навчається на цих даних. Алгоритм проходить через дані та визначає закономірності в них. Наприклад, припустимо, що ми хочемо визначити, чиє обличчя присутнє на даному зображенні, є кілька речей, які ми можемо розглядати як візерунок:

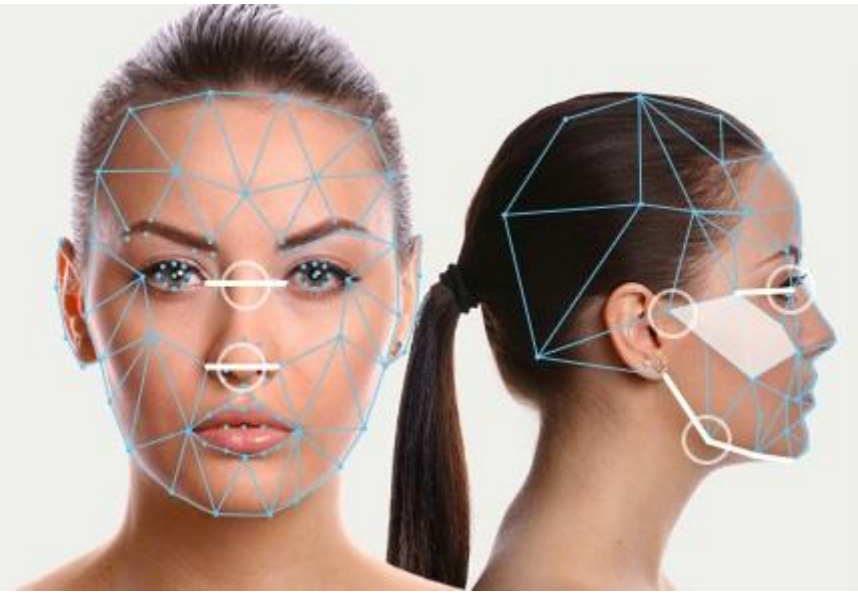


Рисунок 2.2 – схематичне зображення вузлових точок

- Висота / ширина обличчя.

  Висота та ширина можуть бути ненадійними, оскільки зображення можна масштабувати до меншої грані. Однак навіть після масштабування незмінними залишаються співвідношення - відношення висоти обличчя до ширини обличчя не зміниться.
- Колір обличчя.
- Ширина інших частин обличчя, таких як губи, ніс тощо.

Очевидно, що в якості візерунка розглядаються різні грані, що мають різні розміри. Схожі грані мають однакові розміри. Складною частиною є перетворення образу обличчя в комп’ютерну форму - алгоритми машинного навчання працюють з інформацією в цифровому

форматі. Це числове зображення обличчя (або елемента в навчальному наборі) називається вектором ознак. Вектор ознак складається з різних чисел у певному порядку.

Висота обличчя	Ширина обличчя	Середній колір обличчя	Ширина губ	Висота носа
23.1	15.8	(255, 224, 189)	5.2	4.4

Як простий приклад, ми можемо відобразити “обличчя” у вектор об’єкта, який може містити різні об’єкти, такі як:

- Висота обличчя (см);
- Ширина обличчя (см);
- Середній колір обличчя (R, G, B);
- Ширина губ (см);
- Висота носа (см)ю

По суті, отримуючи зображення, ми можемо нанести на карту різні об’єкти та перетворити їх у вектор об’єктів.

Вектор об’єкта Таблиця 2.2

Отже, наше зображення тепер є вектором, який можна представити як (23.1, 15.8, 255, 224, 189, 5.2, 4.4). Звичайно, може бути незліченна кількість інших особливостей, які можна отримати з зображення (наприклад, колір волосся, волосся на обличчі, окуляри тощо). Однак на прикладі розглянемо лише ці 5 простих особливостей.

Тепер, коли ми закодували кожне зображення у вектор об’єкта, проблема стає набагато простішою. Очевидно, що коли ми маємо 2 грані (зображення), які представляють одну і ту ж особу, отримані вектори ознак

будуть досить подібними. Інакше кажучи, "відстань" між двома векторами характеристик буде досить малою.

Машинне навчання може допомогти нам тут у двох задачах:

1. Виведення вектора ознак: складно вручну перерахувати всі об’єкти, оскільки їх дуже багато. Алгоритм машинного навчання може розумно позначити багато таких функцій. Наприклад, складними ознаками можуть бути: співвідношення висоти носа та ширини чола. Зараз людині буде досить важко перерахувати всі такі особливості “другого порядку”.
2. Алгоритми узгодження: Після отримання векторів функцій алгоритм машинного навчання повинен зіставити нове зображення із набором векторів ознак, присутніх у корпусі.

Тепер ми маємо базове розуміння того, як працює розпізнавання обличчя. Розпізнавання обличчя може бути реалізоване або як повністю автоматизована система, або як напівавтоматизована система. У перших випадках втручання людини не потрібно, але в другій частині воно є обов’язковим. Це найкращий метод, який слід використовувати при розгортанні пристрою розпізнавання обличчя.

Виявлення обличчя відбувається спочатку. Алгоритми зазвичай кружляють по різних вікнах, шукаючи обличчя з певним розміром. Усередині цих ящиків система виявляє орієнтири обличчя та призначає оцінку, забезпечуючи рівень впевненості щодо того, чи є зображення обличчям. Після підтвердження наявності обличчя, технологія створює шаблон, який, як правило, базується на таких факторах, як відносна відстань між очима, пляма безпосередньо під носом та над губою, та відстань вуха до вуха. Потім розроблене математичне подання порівнюється з іншими виявленими гранями. Подібність у співвідношенні відстаней між різними точками обличчя, як правило, зосередженими

навколо ознак, таких як ніс, очі, вуха та рот, дає оцінку за логарифмічною шкалою.

Чинники, що впливають на якість оцифровування.

- Середовище камери має важливе значення:

Коли користувач стикається з камерою, що стоїть приблизно на відстані 60см від нього, система визначає обличчя користувача та виконує збіги із заявленою особою або базою даних обличчя. Можливо, користувачеві потрібно буде рухатися та повторити спробу перевірки, виходячи з його положення обличчя. Система зазвичай приймає рішення менш ніж за 5 секунд. Незалежно від того, яка техніка використовується, розпізнавання обличчя працює краще, коли є хороший набір зображень обличчя для роботи. Дуже важливо мати постійне та контрольоване освітлення, роздільну здатність камери, положення обличчя та обмежений рух.

- Освітлення:

Освітлення на обличчі має бути досить яскравим, щоб датчик камери не видавав шуму. Також має бути достатньо світла, щоб забезпечити достатню контрастність для алгоритму розпізнавання. Багато з цих систем вимагають принаймні від 300 до 500 люксів освітлення. Йдеться про світло, яке ми бачимо в звичайному офісному робочому середовищі. Освітлення повинно бути послідовним, щоб тіні не створювали фальшивих чи помилкових зображень обличчя.

- Роздільна здатність:

Роздільна здатність IP-камери залежить від типу використовуваної системи розпізнавання та загального поля зору. Загалом, чим ширше поле зору, тим більше роздільної здатності вам знадобиться. Багато систем вимагають певної мінімальної кількості пікселів для певних рис обличчя.

Наприклад, вам може знадобитися 60 пікселів між двома очима (міжочна відстань). Деякі інші системи вимагають принаймні від 80 до 120 пікселів по обличчю. Коли ми знаємо вимоги до роздільної здатності системи розпізнавання облич, ми можемо розрахувати роздільну здатність камери.

Ось приклад. Якщо ми використовуємо аналітичну систему, яка базується на відстані між двома очима, нам спочатку потрібно знати, якою буде ця відстань. База даних та дослідження були проведені для антропометричного опитування особового складу армії США в 1988 році. У цьому дослідженні середній розмір для чоловіків становив близько 65 мм, а для жінок у середньому близько 62 мм. Найбільша виміряна відстань становила 74 мм, тоді як найкоротша - близько 55 мм. Для розрахунку роздільної здатності IP-камери, необхідної для цього типу розпізнавання обличчя, найкраще використовувати найкоротший розмір 55 мм.

- Нам потрібно 60 пікселів на 55 мм або 60/55 = 1,09 пікселів / мм.
- Далі нам потрібно вирішити, яке поле зору (FOV) ми хотіли б. Припустимо, ми вирішили, що горизонтальне поле зору дорівнює 1524 мм (що становить приблизно 1,5 метри у ширину).
- Для досягнення 1,09 пікселів / мм на 1524 мм FOV нам потрібно 1,09 x 1524, що дорівнює 1663 горизонтальним пікселям.
- Далі ми шукаємо камеру, яка має принаймні таку кількість горизонтальних пікселів. Більше - краще.
- 2-мегапіксельна камера (1920 x 1080) перевищує цю вимогу, тоді як 1-мегапіксельна камера (1280 x 1024) не працює.

Для цього додатка найкраще підходить 2-мегапіксельна IP-камера з 1920 пікселями по горизонталі. Якщо ми хочемо переглянути більшу площу, нам потрібно буде збільшити роздільну здатність камери.

Наприклад, для 3 метрової ширини потрібна роздільна здатність удвічі більша.

Положення обличчя:

Один програмний продукт для розпізнавання обличь був використаний у церквах для визначення того, хто брав участь у службі. Програмне забезпечення Churchix досить добре працює в церкві, де рівень освітлення правильний, усі дивляться вперед і навіть мають однаковий вираз обличчя (більшу частину часу). У багатьох додатках завдання полягає в розміщенні обличчя в правильному напрямку. Найкраще встановити систему розпізнавання обличь у дверному отворі або на воротах, де, ймовірно, всі спрямовані у правильному напрямку, і вони наближаються до камери по кілька людей одночасно. Є деякі системи, які можна використовувати в менш обмежених середовищах, таких як натовп. Задача розпізнавання обличчя в натовпі складніша, за біометричну ідентифікацію обличчя.

Обмежений рух:

Системи камер повинні мати можливість захопити обличчя, і якщо руху буде занадто багато, це може призвести до поганого відеозображення. Для підвищення продуктивності, коли люди швидко рухаються, для систем потрібні камери, які можуть підтримувати високу частоту кадрів. Зазвичай достатньо 30 кадрів в секунду, але якщо очікується більше руху, може знадобитися камера з швидкістю захоплення зображення 60 кадрів в секунду. Алгоритмам розпізнавання обличчя потрібен певний час для обробки даних, тому, якщо через систему протікає занадто багато людей, процес розпізнавання може бути недостатньо швидким. По мірі збільшення кількості виявлень необхідні більш ефективні апаратні ресурси.

Весь процес розпізнавання обличчя відеокамерою проходить кілька етапів (рис. 2.3):

- Система приймає відеопотік і аналізує всі кадри в режимі реального часу. Коли обличчя виявлено, алгоритм робить кілька його знімків і вирізає обличчя з них.
- У реальному житті людина може рухатися, повертатися і навіть опускати голову. Отже, на наступному етапі система вирівнює грані на кожному зображенні так, щоб алгоритм на наступному етапі міг аналізувати обличчя спереду для вищої точності.
- Після цього система створює унікальний вектор, який складається з опису рис обличчя (наприклад - відстань між очима, довжина чола, носа, тон шкіри тощо). Однак фактичний процес створення векторів набагато складніший, але дуже схожий на спосіб, яким ми помічаємо риси обличь. Вектор - це не проста таблиця параметрів; швидше, це складна «ідея» того, що визначає це конкретне обличчя і чим воно відрізняється від інших, перетворене у вектор даних за допомогою певного математичного процесу. Цей вектор створюється для кожного кадру з цим обличчям. В результаті ми маємо купу векторів.
- Система їх аналізує множину векторів, кластеризуючи та витягаючи ключові вектори, які будуть відправлені до сховища. Далі ця множина векторів порівнюється з іншими наявними у базі.
- Потім компаратор надсилає результати через API до будь-якої іншої системи.

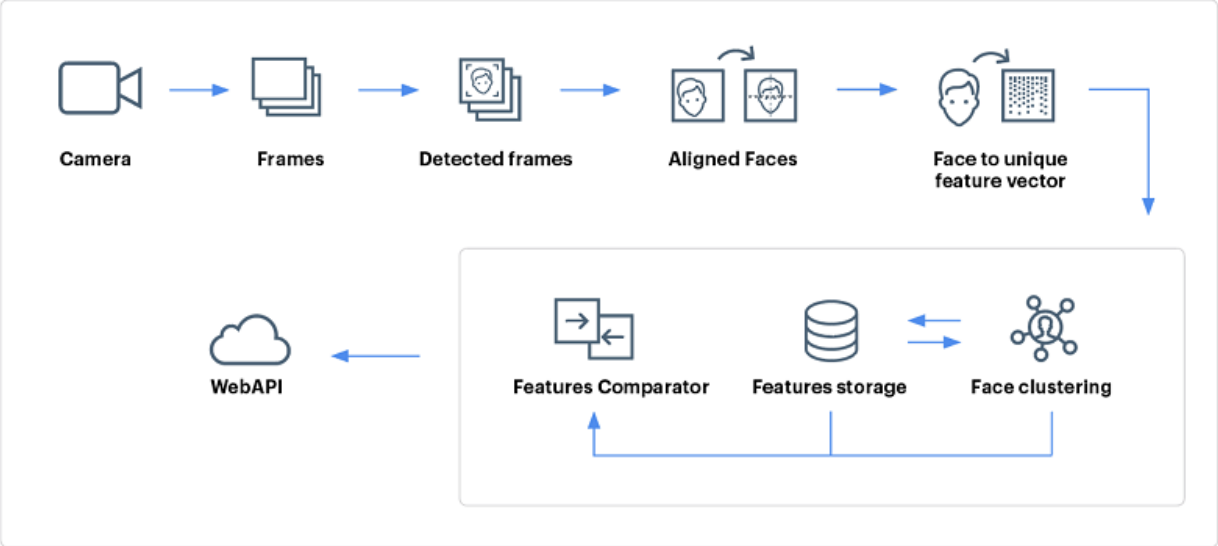


Рисунок 2.3 – процес розпізнавання обличчя

Порівняльний аналіз існуючих архітектур компаратора:

В даному контексті, перспективними є наступні методи розпізнавання обличчя:

1. Геометричний (Шаблонний) метод :

Алгоритми розпізнавання обличчя, класифіковані як алгоритми на основі геометрії або шаблонні алгоритми. Методи, засновані на шаблонах, можуть бути побудовані з використанням статистичних інструментів, таких як SVM(Support Vector Machines), PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), методи ядра або перетворення трасування. Методи, засновані на геометричних ознаках, аналізують локальні риси обличчя та їх геометричні взаємозв'язки. Також відомий як функціональний метод.

2. Піксельний (Цілісний) метод:

Даний метод використовує за основні параметри співвідношення між елементами або зв'язок функції з цілим обличчям, багато дослідників дотримувалися цього підходу, намагаючись вивести найбільш відповідні характеристики. Деякі методи використовують взаємозв'язки та відстані між основними параметрами обличчя. Деякі методи ПММ також потрапляють до цієї категорії.

3. Модельний метод (метод зовнішнього вигляду):

Метод, що базується на зовнішньому вигляді, показує обличчя відносно кількох зображень. Зображення розглядається як вектор великого розміру. Цей прийом зазвичай використовується для виведення простору об’єктів із поділу зображення. Зразок зображення порівнюється з навчальним набором. З іншого боку, підхід, заснований на моделі, намагається змоделювати обличчя. Новий зразок, реалізований у моделі, і параметри моделі, використовувані для розпізнавання зображення.

Метод, заснований на зовнішньому вигляді, можна класифікувати як лінійний або нелінійний. Ex-PCA, LDA, IDA використовуються в прямому підході, тоді як ядро PCA використовується в нелінійному підході. З іншого боку, метод, заснований на моделі, може бути класифікований як 2D або 3D Ex-Elastic Bunch Graph Matching.

4. Метод НМ (статистичний метод):

4.1. Підбір шаблону

При застосуванні методу підбору шаблону вхідні набори подаються у вигляді зразків моделей, пікселів, текстур. Функція розпізнавання зазвичай є кореляцією або мірою відстані.

4.2. Статистичний підхід

При використанні статистичного підходу, закономірності виражаються у вигляді ознак. Мета полягає у виборі та застосуванні правильного статистичного інструменту для вилучення та аналізу.

Є багато статистичних інструментів, які використовуються для розпізнавання обличчя. Ці аналітичні засоби використовуються у наступних методах класифікації:

4.2.1. Principal Component Analysis

Одним із найбільш часто використовуваних статистичних методів є аналіз основних компонентів. Математична процедура виконує зменшення

розмірності шляхом вилучення основної складової багатовимірних даних (рис. 2.4)

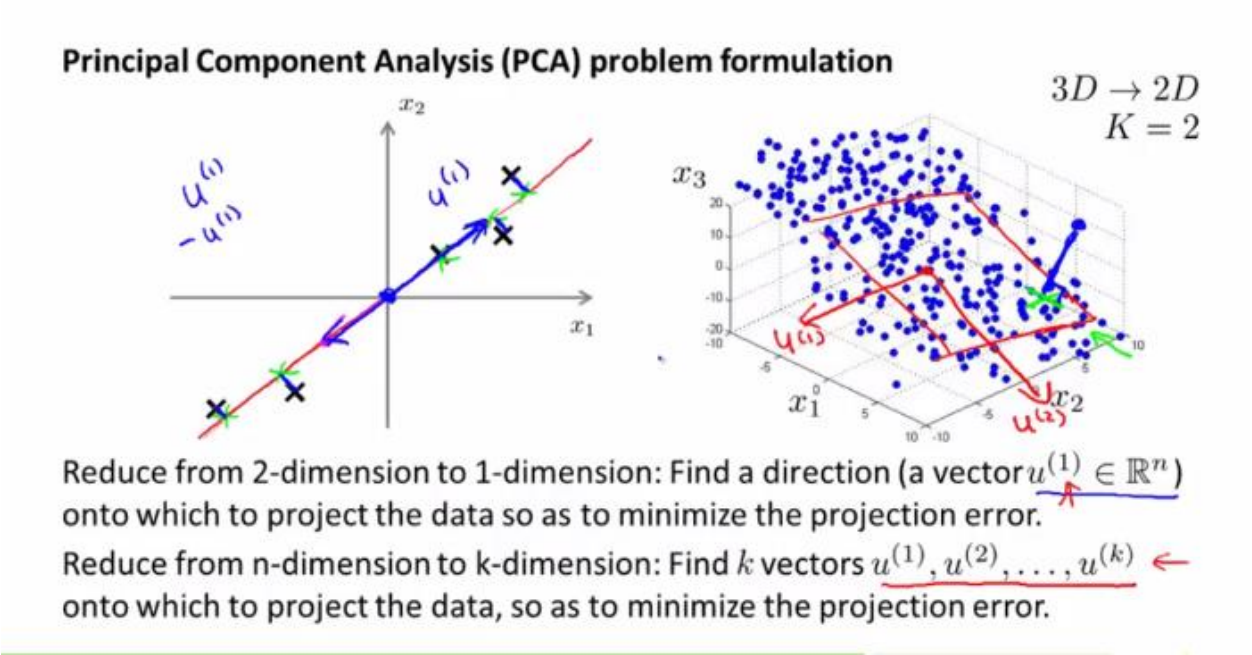
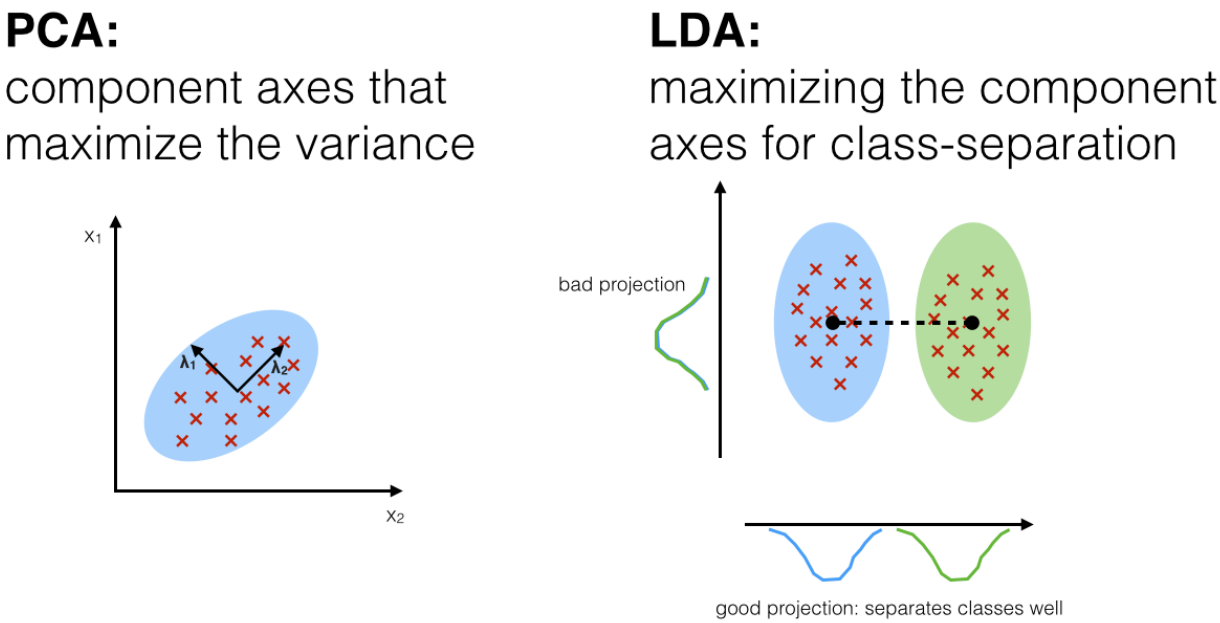


Рисунок 2.4 – двовимірна апроксимація тривимірного зразка за допомогою PCA

4.2.2. Discrete Cosine Transform

Дискретне косинусне перетворення засноване на дискретному перетворенні Фур'є, і, отже, шляхом ущільнення варіацій воно може бути використане для перетворення зображень та дозволяє ефективно



зменшувати розмірність.

Рисунок 2.5 – порівняння методів PCA та LDA

4.2.3. Linear Discriminant Analysis

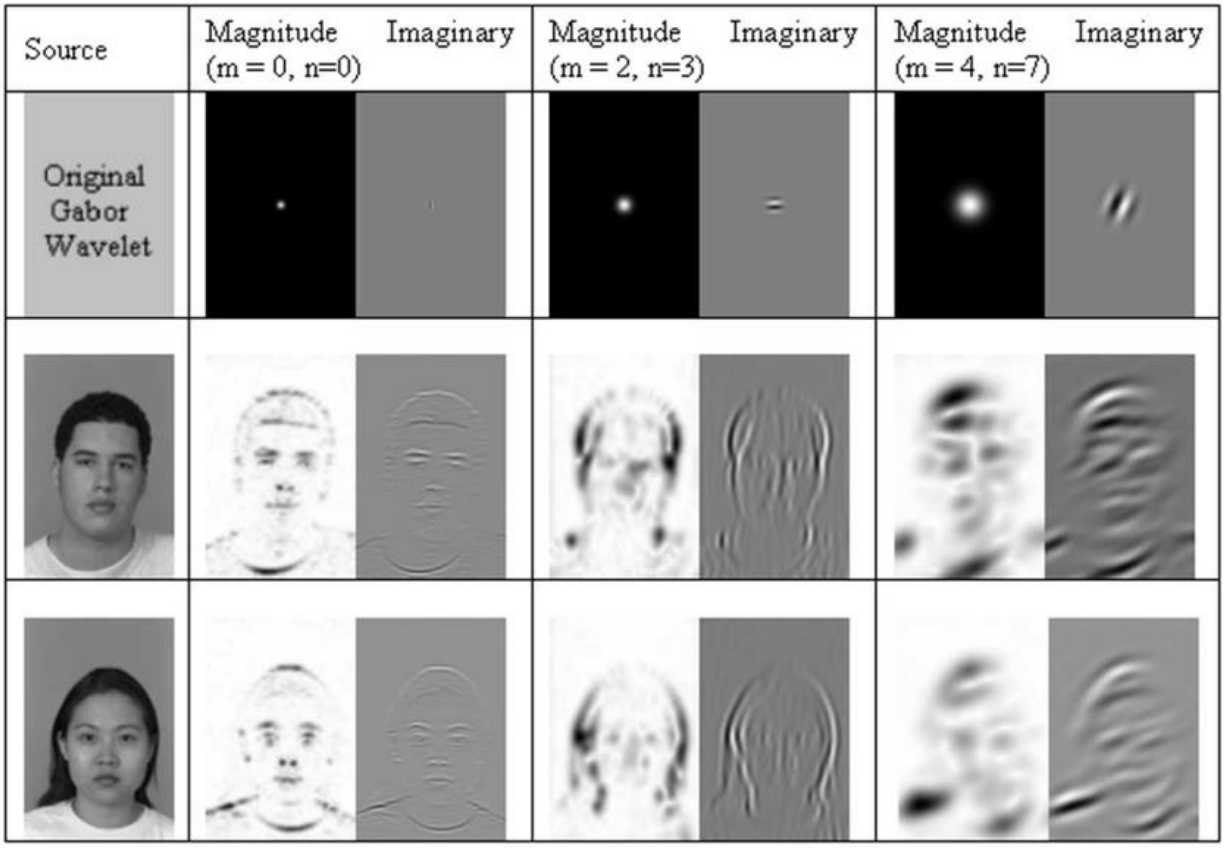
LDA широко використовується для пошуку лінійної комбінації функцій при збереженні сегрегації класів. На відміну від PCA, LDA намагається моделювати різницю між рівнями. Для кожного рівня LDA отримує різницю в кількох проекційних векторах (рис.2.5).

4.2.4. Locality Preserving Projections

Це найкраща альтернатива PCA для збереження локальної структури та проектування. Алгоритми розпізнавання шаблонів зазвичай шукають найближчий шаблон або сусідів. Таким чином, структура, що підтримує можливість застосування LLP, може пришвидшити визначення.

4.2.5. Gabor Wavelet

Даний метод передбачає застосування двовимірних вейвлетів Габора, завдання яких полягає у моделюванні нейрофізіологічних перетворень що відбуваються в мозку ссавців при передачі візуального



зображення до зорової кори головного мозку. Функції Габора, запропоновані Доугманом, є локальними просторовими смуговими фільтрами, які досягають теоретичного обмеження для спільної роздільної здатності інформації в 2D просторовій та 2D областях Фур'є (рис. 2.6).

Рисунок 2.6 – вейвлет Габора в різній величині

4.2.6. Independent Component Analysis

Мета ICA полягає у забезпеченні незалежного, а не некорельованого подання зображення. ICA є альтернативою PCA, що забезпечує більш потужне представлення даних. Це дискримінантний критерій аналізу, який може бути використаний для посилення PCA.

4.2.7. Kernel PCA

Основна методологія полягає у застосуванні нелінійного відображення до вхідних даних, та вирішенні лінійного PCA у отриманому підпросторі ознак.

4.3. Нейронні мережі

НМ використовують для розпізнавання та класифікації шаблонів. Кохонен першим показав, що нейронну мережу можна використовувати для розпізнавання вирівняних та нормалізованих граней. Існують методи, які виконують вилучення особливостей за допомогою нейронних мереж. Існує багато методів, які поєднують з такими інструментами, як PCA або LCA, і створюють гібридний класифікатор для розпізнавання обличь. Наприклад, Feed Forward NN, Self-Organizing Maps with PCA, CNN with multi-layer perception, що може підвищити ефективність моделей (рис. 2.7).

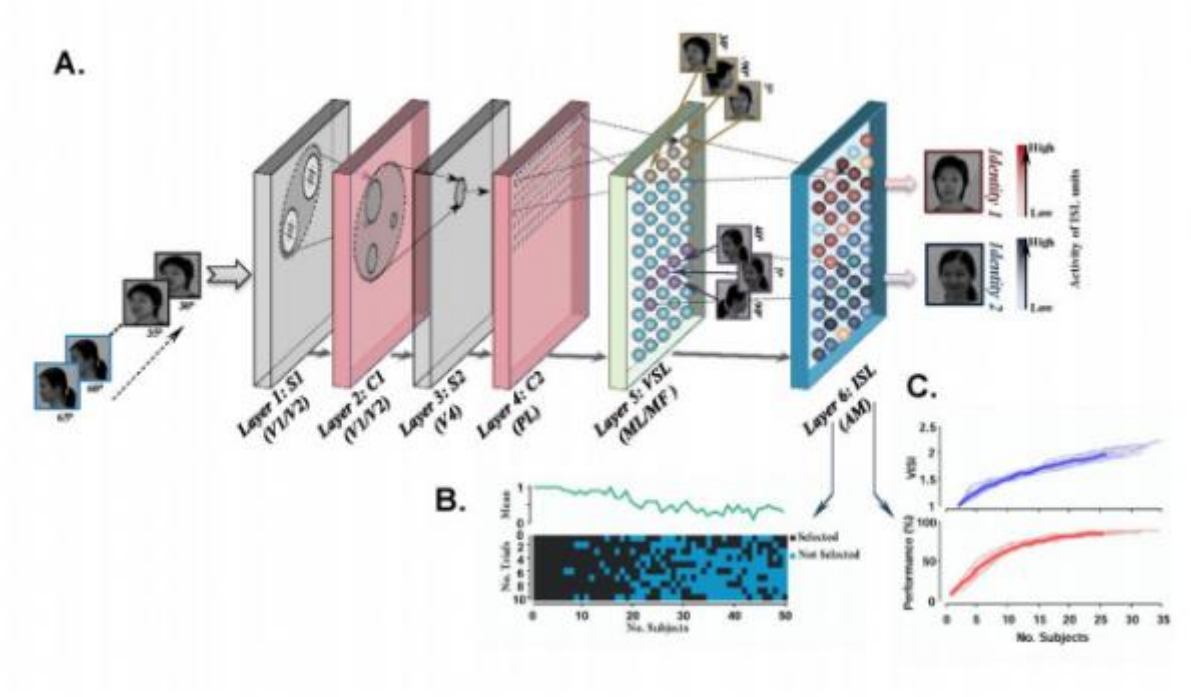


Рисунок 2.7 – архітектура DNN для розпізнавання обличчя

4.3.1. НМ та фільтри Габора

Алгоритм досягає розпізнавання обличчя шляхом реалізації багатошарового персептрона з алгоритмом зворотного поширення. По-перше, є крок попередньої обробки. Кожне зображення нормалізується на фазах контрасту та освітлення. Потім кожне зображення обробляється через фільтр Габора. Фільтр Габора має п'ять параметрів орієнтації та три просторові частоти, тому існує 15 довжин хвиль Габора (рис. 2.8).

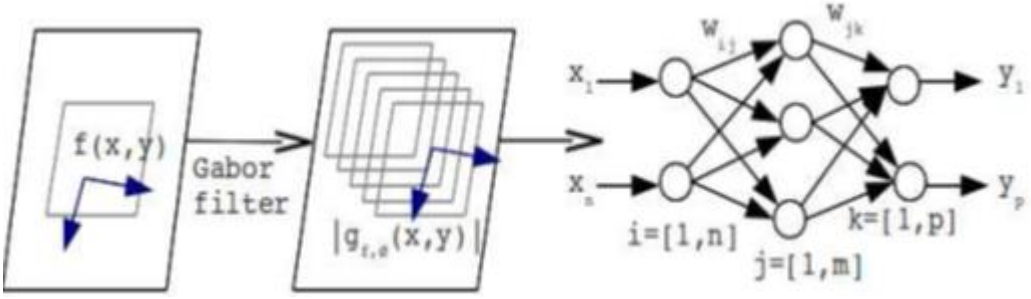
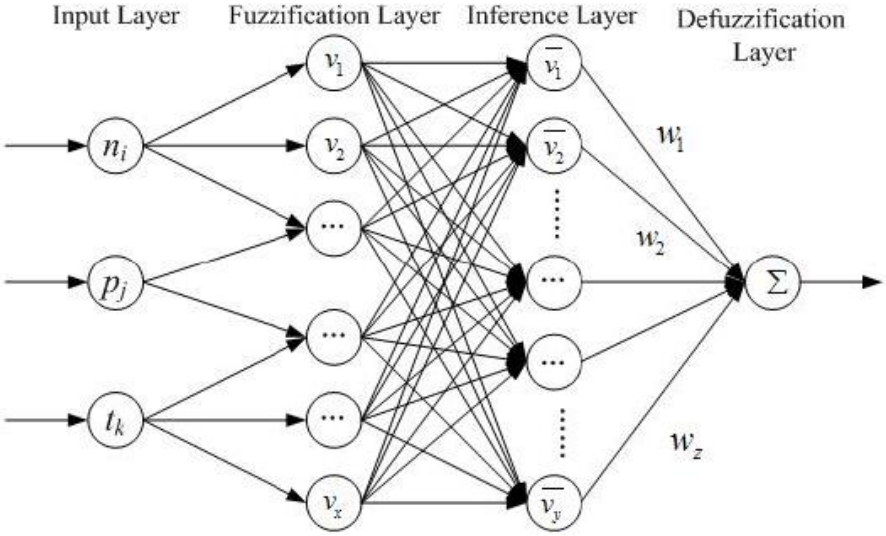


Рисунок 2.8 – застосування НМ з фільтрами Габора

4.3.2. НМ з ПММ

Приховані моделі Маркова - це статистичний інструмент, який використовується для розпізнавання обличь в парі з нейронними мережами. Вхідні дані цього 2D HMM-процесу є надходять з штучної нейронної мережі, що забезпечує алгоритму належне зменшення розмірності.



4.3.3. Нечіткі НМ

Рисунок 2.9 – нечітка НМ

Нечіткі нейронні мережі для розпізнавання обличчя використовуються з 2009 року. У цій системі розпізнавання обличь використовується багатошаровий персептрон. Концепція цього підходу полягає в тому, щоб зафіксувати поверхні прийняття рішень у нелінійних наборах завдання, яке простий MLP навряд чи може виконати. Вектори характеристик отримані за допомогою перетворень довжини хвилі Габора (рис. 2.9).

В ході аналізу була встановлена перспективність застосування згорткових нейронних мереж (CNN) для реалізації механізму компаратора. Рисунок 2.10 зображує алгоритм розпізнавання згідно обраної архітектури.

Фінальним кроком є побудова математичної моделі для навчання і подальшого тестування.

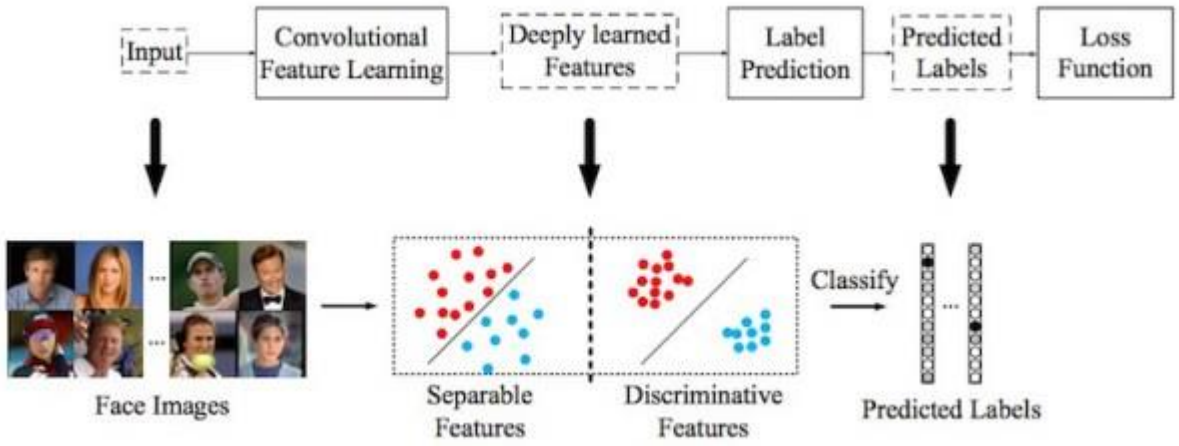
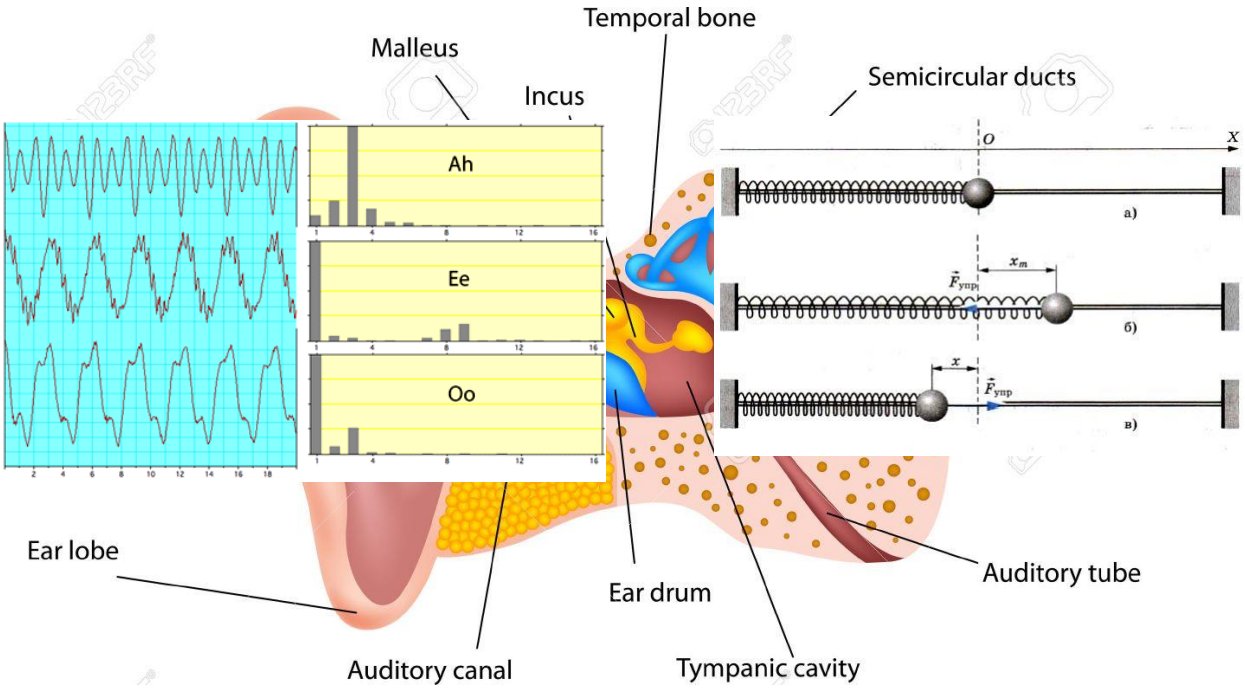


Рисунок 2.10 – алгоритм візуальної ідентифікації на основі архітектури CNN

2.3. Визначення перспективних методів голосової біометричної ідентифікації

Звук – це фізичне коливання, вібрація, що має бути трансформована в комп'ютерну форму для подальшого аналізу. Людське вухо здатне до такої трансформації за своєю природою, таким чином, що різні складові органу реагують на окремі частотні діапазони і виконують нелінійне перетворення з вхідними звуками. Таким чином, людське вухо наділене приймачами для окремих частотних діапазонів, що видно з рис. 2.11. Рисунки 2.12 та 2.13 зображують взаємозв'язок між осциляторами, що створюють різні частоти, та окремими тонами.

Рисунок 2.11 – Внутрішня будова вуха



Згідно перетворенню Фур'є, кожен звук є комбінацією синусоїдальних хвиль різних параметрів.

Рисунок 2.12 – Частотні залежності

Рисунок 2.13 – Механічні коливання

Синусоїдальні хвилі різної амплітуди і частоти сприймаються людиною як звуки різних тонів. При комбінації хвиль, їх результуюча резонує, утворюючи піки, або ями в частотному діапазоні. Ця особливість надає можливість сприйняття інформації шляхом диференціації

частотного діапазону, і аналізу його змістовної частини. Дані умови повинні бути реалізовані в розробленій моделі розпізнавання:

- Диференціація та аналіз частотного діапазону 40-600Гц;
- Диференціація сигналу на набори гармонік.

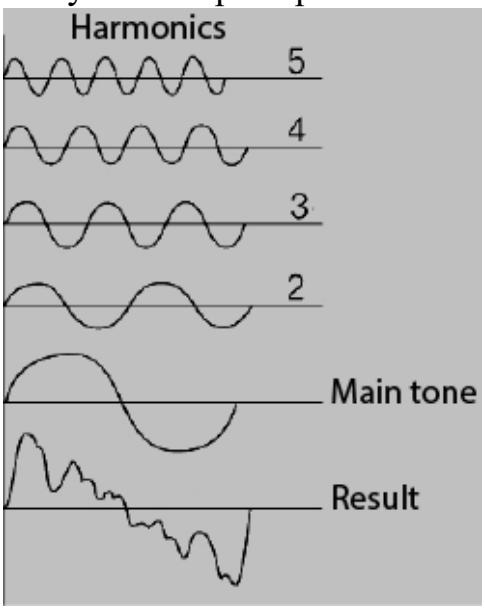


Рисунок 2.14 – Гармонічна диференціація сигналу

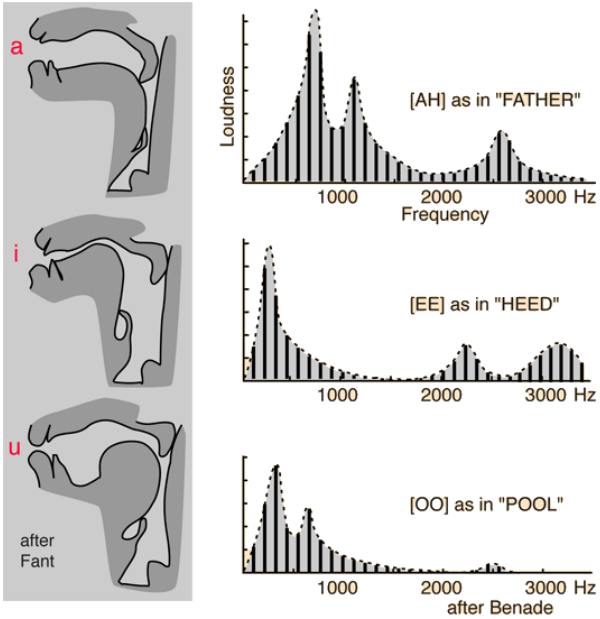


Рисунок 2.14 – Частотний розподіл між звуками

На рисунку 2.14 зображені набори гармонік людського голосу, що при накладанні утворюють результуючий сигнал. Рисунок 2.15 зображує зв’язок між різними типами звуків та їх частотними характеристикам.

Отже, розроблена модель повинна надавати можливість поділу результуючого сигналу на гармоніки, набір яких є індивідуальним біометричним показником, і дозволяє ідентифікувати суб'єкта.

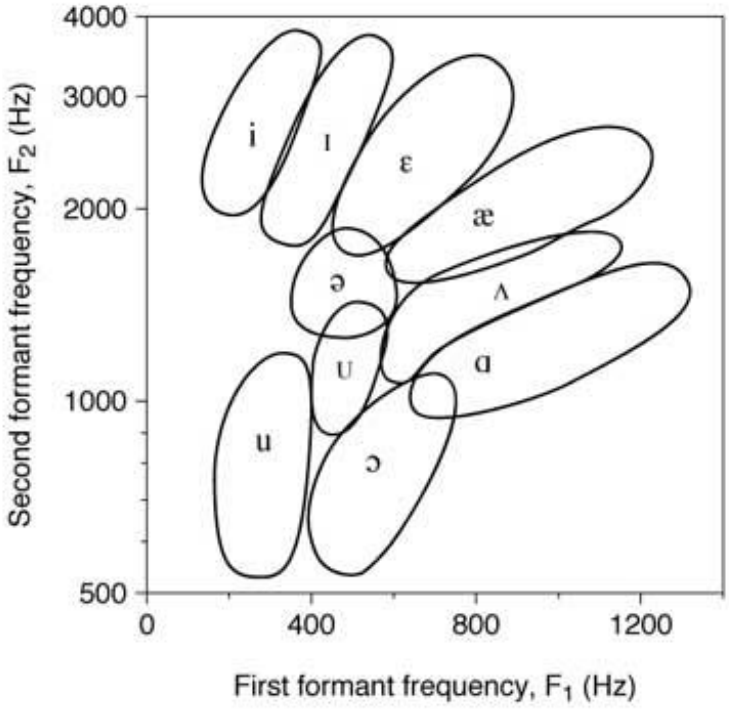


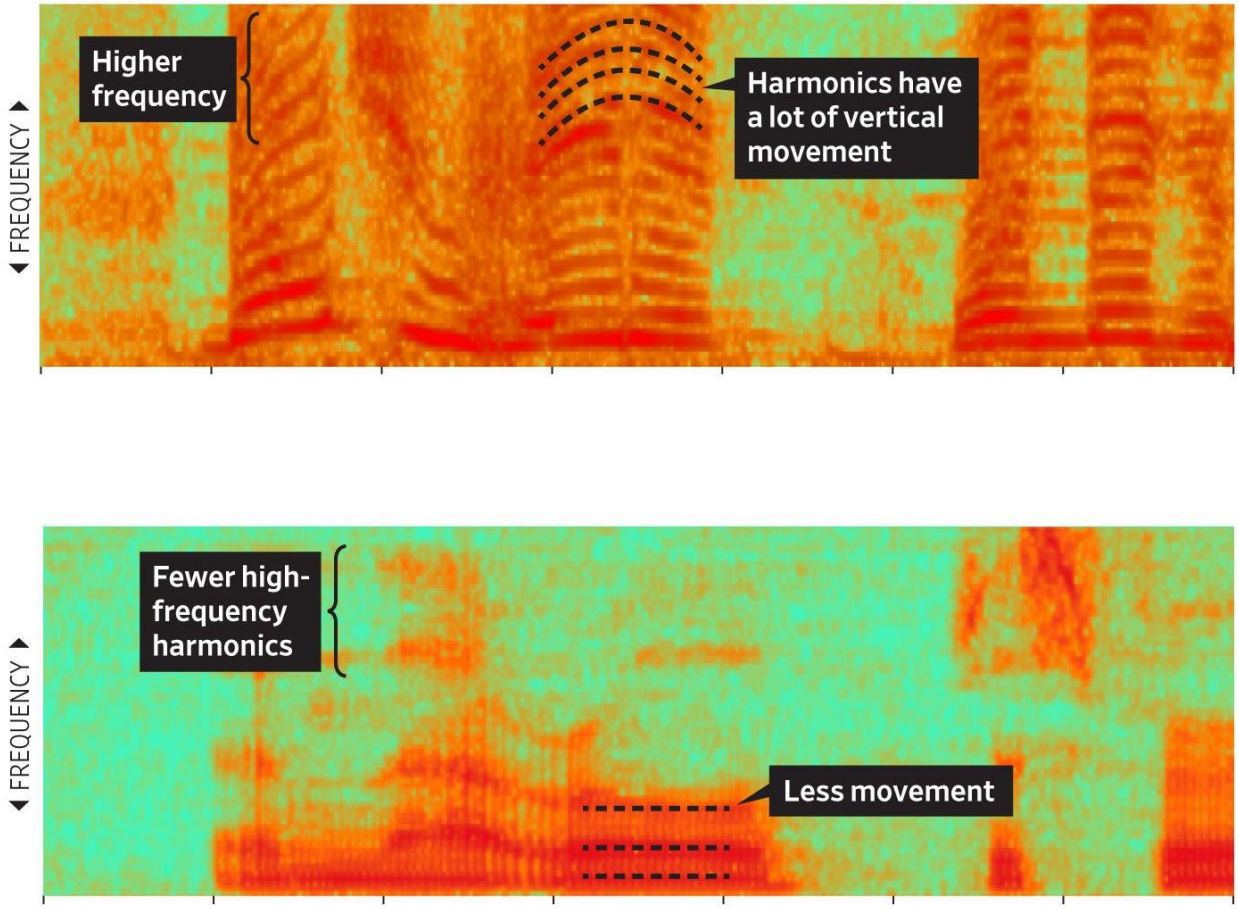
Рисунок 2.16 – Розподіл голосних звуків на частотному діапазоні

Згідно рис. 2.16, голосні звуки займають велику частину діапазону, що спрощує їх ідентифікацію порівняно до приголосних. Цей фактор, свого часу, визначив шлях розвитку технології розпізнавання мовлення - перші моделі розпізнавання були побудовані за принципом аналізу послідовності голосних у слові.

Так як звук є форматом даних з динамічними показниками, розроблений алгоритм повинен аналізувати частотний спектр в окремі моменти часу. Рисунок 2.17 зображує залежність між частотними показниками сигналу та часом. Дослідження цих показників дозволяє відокремлювати фонеми, для подальшого складання в набори та порівняння з словниками моделі і як наслідок розпізнавання. Людина наділена здатністю до аналізу аналогового сигналу, отже розроблена модель повинна мати схожу структуру і мати можливість інтерпретації вхідного сигналу в дискретній формі, яка здатна бути проаналізована моделлю. Рисунок 2.18 зображує

загальний стек методів, що застосовуються для трансформації голосового сигналу.

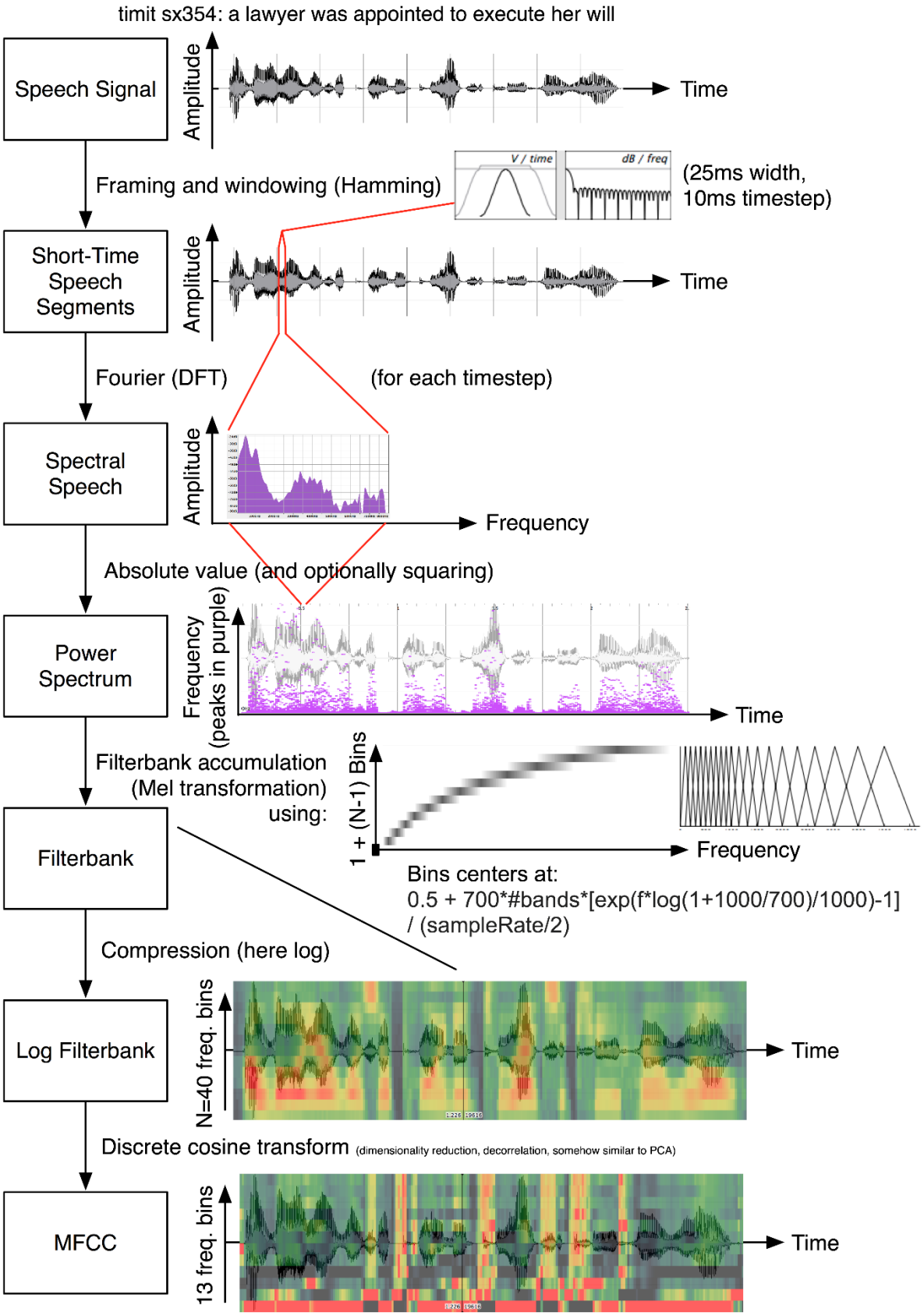
Рисунок 2.17 – Залежність між амплітудою частоти та часом



Алгоритми що застосовані у стеку:

1. Вхідний аналоговий сигнал є залежністю амплітуди від часу;
2. Вхідний сигнал розкладається на короткі сегменти довжиною 10-25 мс – фрейми, шляхом дискретизації аналогового сигналу. Проводиться операцію фреймінгу, при якій дискретні частини множаться на функцію, що прибирає неінформативні частотні діапазони;
3. Застосовується ДПФ що подає фрейми у вигляді амплітудно частотної характеристики;
4. Фрейми формують у динамічну послідовність спектральних залежностей. Вихідна послідовність зображує співвідношення між амплітудою результуючої частоти фрейму та часом;

5. Відбувається фільтрація сигналу що має назву Мел-трансформація. Фільтр обрізає сигнал до діапазону 50Гц з найбільшою



інформативністю;

Рисунок 2.18 – алгоритм трансформації голосового сигналу

6. Компресія сигналу зменшує величину результуючих частот для спрощення подальшого аналізу;
7. Застосування алгоритму MFCC, який стискає сигнал до 13 формуючих частот. Алгоритм DCT в свою чергу займається пошуком повторюваних гармонік з метою їх видалення та скорочення кількості результуючих гармонік до 13.

Вихідна інформація подана в даній формі готова до аналізу НМ. У випадку розпізнавання частотних та тембральних характеристик голосового сигналу, заради ідентифікації суб'єктів, доцільним є використання методу розпізнавання фонем. В даному випадку, процес розпізнавання зводиться до порівняння фонем з існуючими словниками – датасетами. Збільшення розмірності словників є основним інструментом збільшення точності розпізнавання.

Першими розробленими датасетами були:

- TIMIT Database (англ.) – 3 години, 1993 рік;
- 199CMUDict (англ.) – 100 тис. слів, 1993.

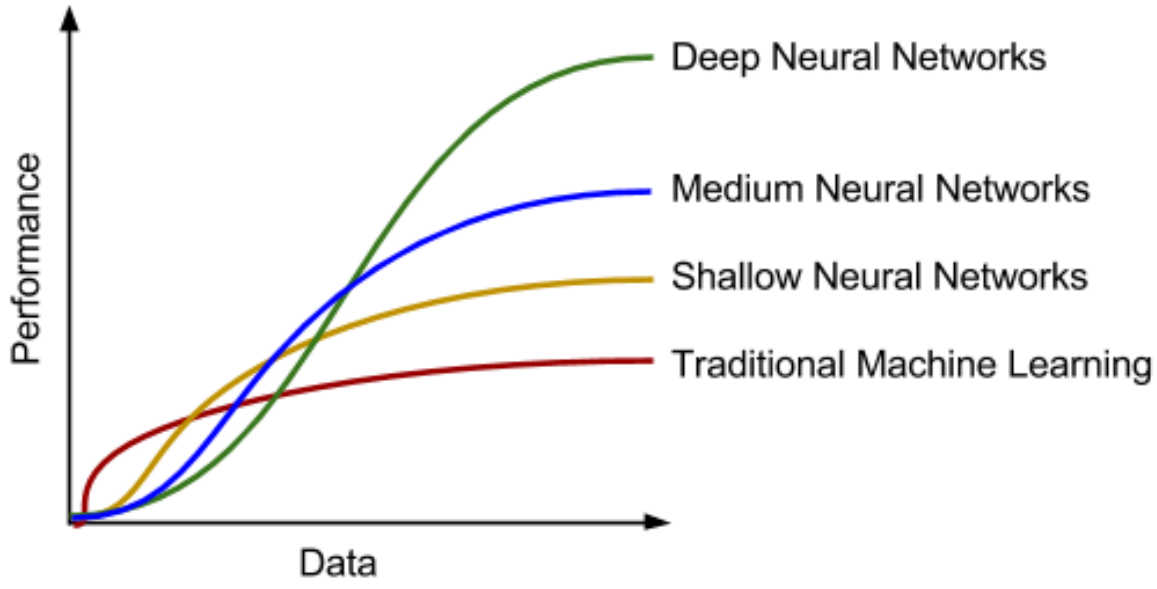


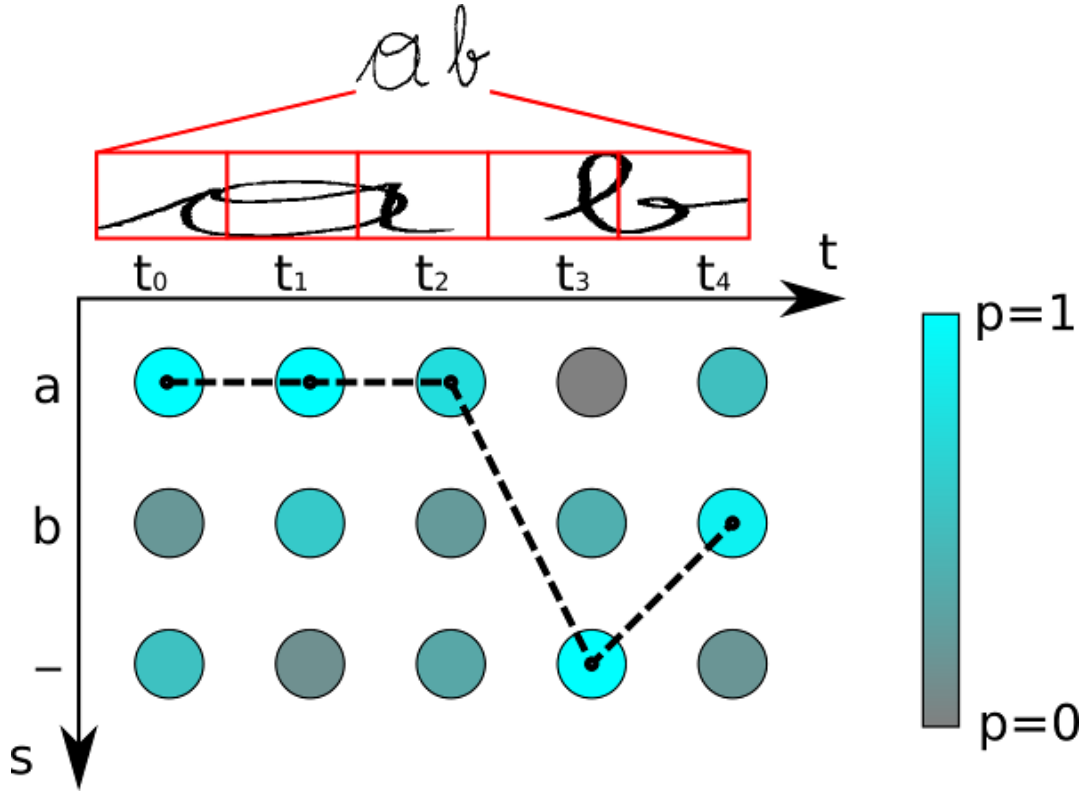
Рисунок 2.19 – Продуктивність алгоритмів розпізнавання мовлення

Згідно рис. 2.19 – продуктивність розпізнавання залежить від алгоритму НМ.

Так як відомі архітектури розпізнавання базуються на аналізі мовлення, слід дослідити відомі рішення, з метою імплементації в фінальну розробку.

По аналогії з розпізнаванням фонем, можливе розпізнавання букв із вхідного сигналу. Використання даного методу примушує до збільшення розмірів словника та потужності системи. Перші словники на основі буквенного аналізу:

- WSJ (англ.) - 81 годин, 1993 ;
- Switchboard (англ.) - 240 годин, 1993;



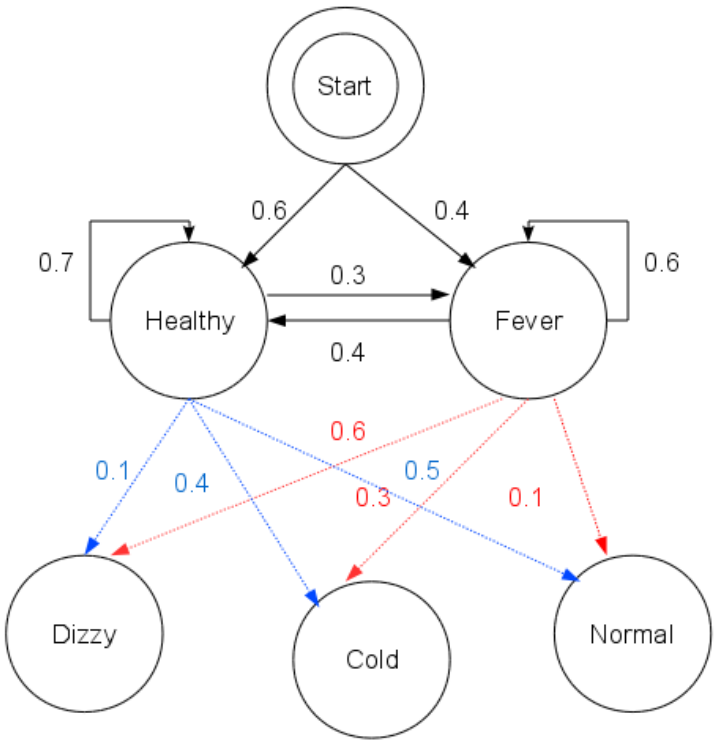
- VoxForge (рос.) - 17 годин, 2009;
- **Fisher (англ.) - 2000 годин, 2004;**
- **LibriSpeech (англ.) - 960 годин, 2015;**
- **Open_TTS (рос.) - більше 3000 годин, 2019.**

Рисунок 2.20 – Алгоритм розмітки (alignment)

Навідміну від перших трьох словників, дані останніх трьох датасетів поділені на слова. Це означає, що для роботи з першими датасетами необхідно підготувати дані, шляхом застосування процесу розмітки (alignment). Відомо декілька алгоритмів розмітки – алгоритм прямого-зворотного зв'язку (Forward-Backward), та алгоритм Вітербі. Зміст вирівнювання полягає у складанні ймовірнісної таблиці, де у відповідність кожній букві алфавіту, або кожній фонемі датасету ставиться ймовірність її знаходження у дискретному кванті часу вхідного сигналу, що зображено на рис. 2.20.

Альтернативою є використання методу HMM, але його ефективність порівняно нижча за Forward-Backward через неможливість опрацювання послідовностей більше 4 звуків, фонем, букв.

Схематичне зображення роботи алгоритму HMM наведене на рис. 2.21. Через недолік у роботі HMM перспективною є ідея реалізації моделі



на основі алгоритму Forward-Backward, який формує ймовірнісну матрицю відповідності оцифрованих дискретних звуків до букв алфавіту (рис. 2.22). Після процесу розмітки, алгоритм знаходить найкоротший шлях за матрицею ймовірностей.

Рисунок 2.21 – Схема роботи HMM

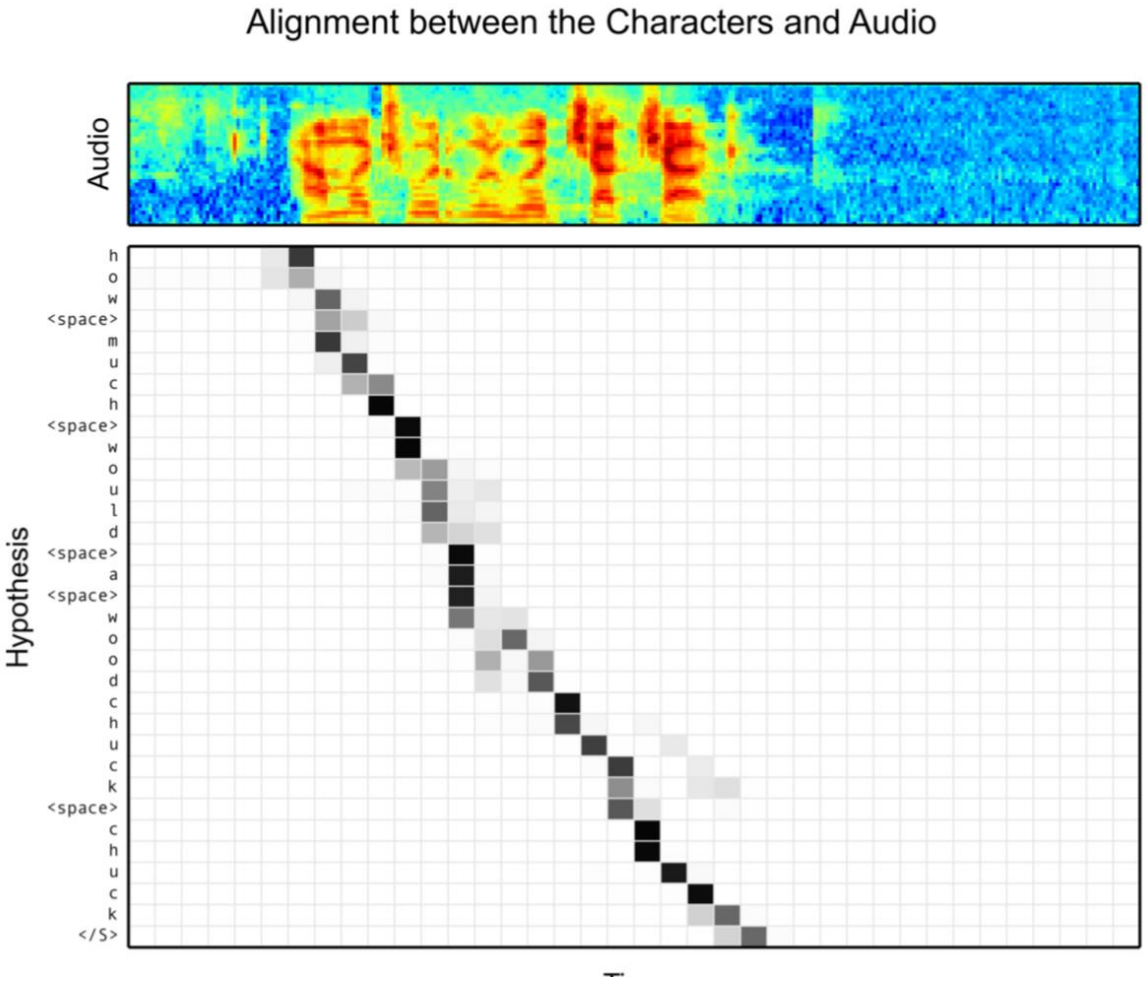


Рисунок 2.22 – матриця ймовірностей алгоритму розмітки

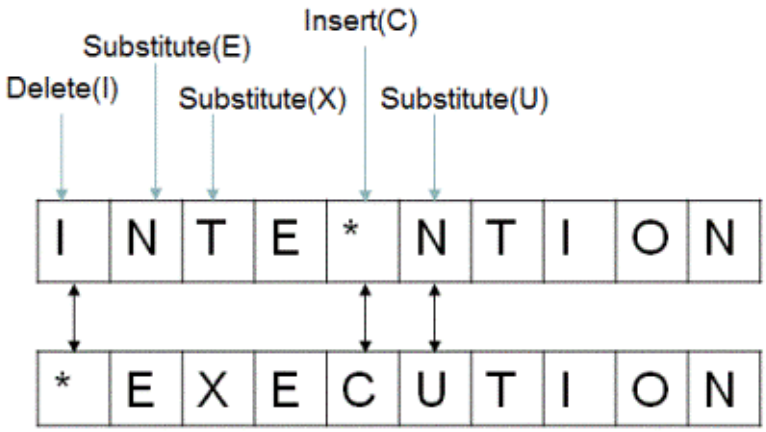
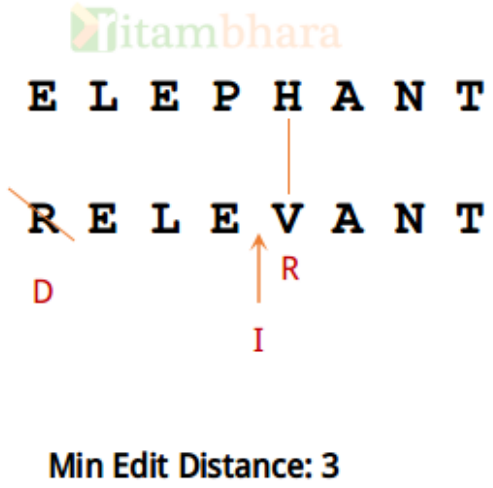


Рисунок 2.23 – Порівняння за допомогою CTC

Алгоритм CTC, що описаний на рис. 2.23-2.25, вимагає застосування Forward-Backward, що порівнює вихідне слово передбачене НМ, з

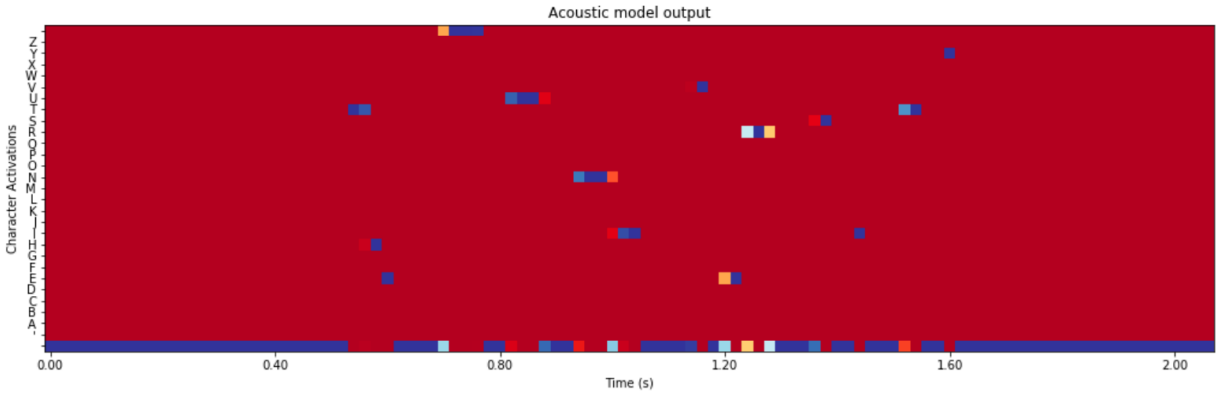
		E	L	E	P	H	A	N	T
	0	1	2	3	4	5	6	7	8
R	1	1	2	3	4	5	6	7	8
E	2	1	2	2	3	4	5	6	7
L	3	2	1	2	3	4	5	6	7
E	4	3	2	1	2	3	4	5	6
V	5	4	3	2	2	3	4	5	6
A	6	5	4	3	3	3	3	4	5
N	7	6	5	4	4	4	4	3	4
T	8	7	6	5	5	5	5	4	3



наявними в датасеті словами. Наступним кроком є знаходження edit distance - цифра в клітинках таблиці, що відповідає кількості перетворень, необхідних для формування вихідного слова. При помилковому визначенні edit distance система отримує штрафні бали, що змінюють показники розпізнавання НМ, ця особливість надає можливість самонавчання НМ.

Рисунок 2.24 – Матриця відстаней алгоритму CTC

Алгоритм CTC для розпізнавання голосу базується на архітектурі PHM - це тип HM, елементи якого утворюють орієнтований граф під час процесу аналізу, що надає можливість обробки динамічних послідовностей:голосу, ручного письма.




```
                    THE  ___ U   N I    VE R   S  I  T Y
Error:                    IRR                          III EEE SSS
Decoder: THE UNIVERSITY
Correct: THEIR UNIVERSITIES
```

Рисунок 2.25 – результат розпізнавання CTC

Застосування CTC вимагає імплементації технологій GRU та LSTM:

1. LSTM – штучна HM, що побудована на основі LSTM модулів. LSTM модуль – повторюваний HM модуль, що здатний запам’ятовувати значення ан певні інтервали часу.
2. GRU (Gated Recurrent Module) = вентильний рекурентний модуль, технологія, що базується на архітектурі LSTM, створена для забезпечення логіки в HM.

Рис. 2.26 схематично зображує архітектуру CTC побудовану на основі даних технологій.

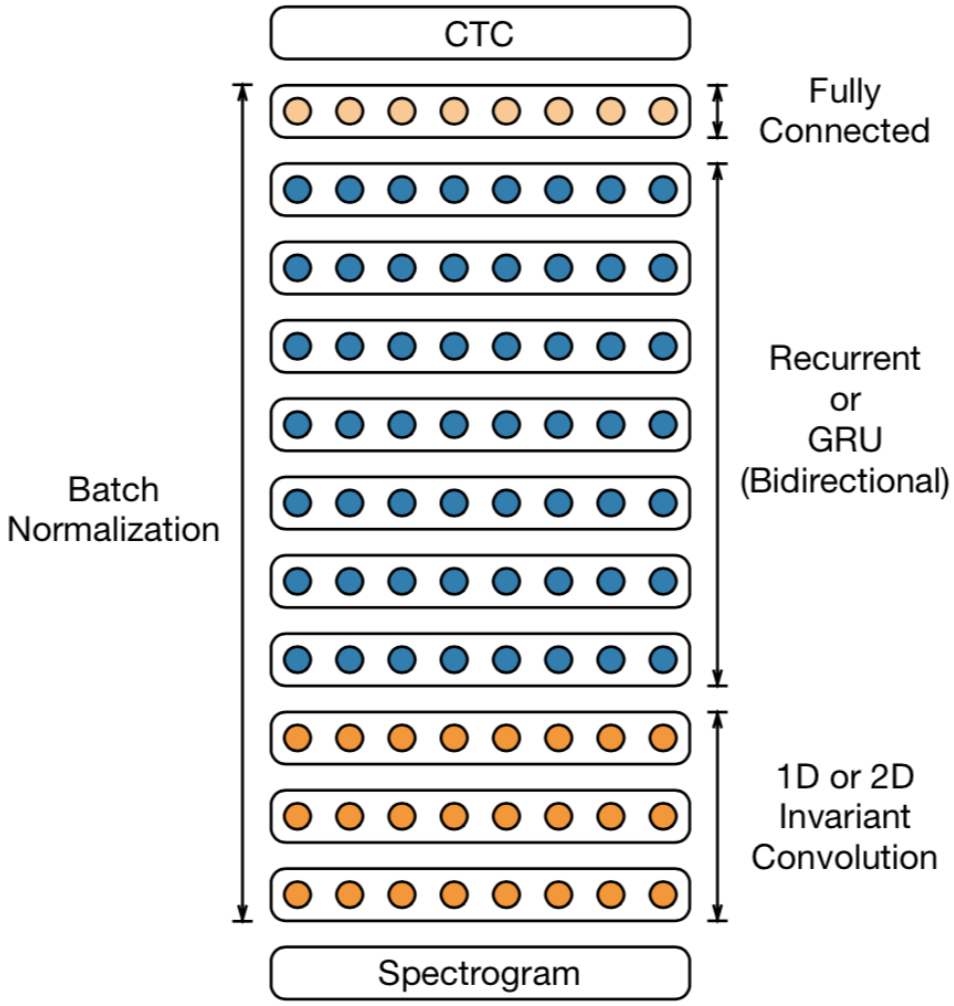


Рисунок 2.26 – Архітектура CTC

На вхід подається спектрограма, яка спочатку представлена в якості одно, або двовимірної структури, що готова до зміни за допомогою шарів CNN. Надалі дані представлені в RNN та GRU у вигляді двонаправленої структури - готові до аналізу динамічні послідовності. Процес нормалізації закінчується перенесенням з'єднання на повністю з'єднаний шар. Даний метод комбінації технологій RNN, GRU та CNN зображено на рис.2.27:

- Одно та двовимірні моделі аналізуються згортковими шарами CNN $h_t^{(1)} - h_t^{(3)}$;
- $h_t^{(f)} - h_t^{(b)}$- двонаправлені шари;
- $h_t^{(5)}$- повністю зв’язаний нейронний шар.

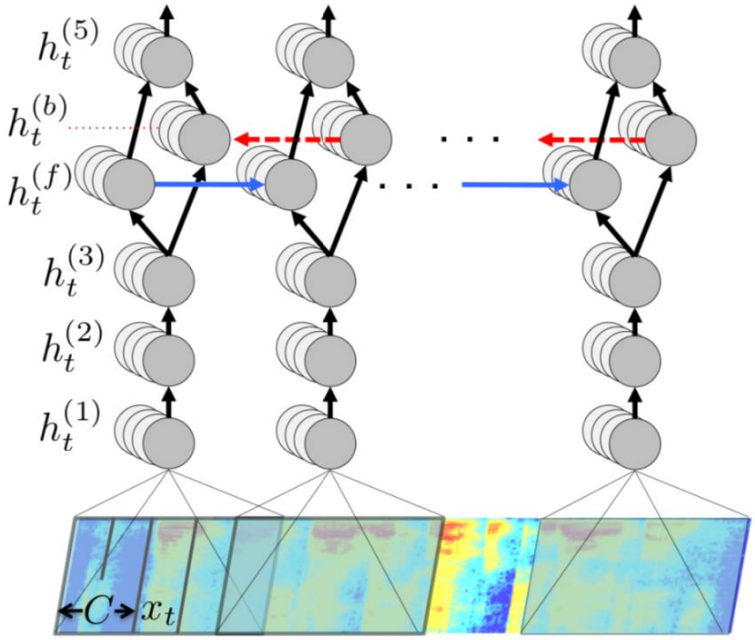


Рисунок 2.27 – Структура PHM

Вихідна інформація зображує сигнал у вигляді набору ймовірностей. Наступним кроком є впровадження системи навчання, що зменшують WER. Здатність самонавчання реалізується шляхом надання достатньої апаратної бази у вигляді процесора з високою тактовою частотою роботи, а для забезпечення швидшої обробки результатів, графічних прискорювачів. Також час тренування варіюється в залежності від конфігурації системи, та розміру датасету.

Architecture	Hidden Units	Train		Dev	
		Baseline	BatchNorm	Baseline	BatchNorm
1 RNN, 5 total	2400	10.55	11.99	13.55	14.40
3 RNN, 5 total	1880	9.55	8.29	11.61	10.56
5 RNN, 7 total	1510	8.59	7.61	10.77	9.78
7 RNN, 9 total	1280	8.76	7.68	10.83	9.52

Рисунок 2.28 – конфігурації архітектур PHM

Рисунок 2.28 відображує результати аналізу окремих конфігурацій архітектур РНМ. Найкращі результати досягнуті на двовимірних згортках, з 7 рекурентних шарів, 1280 прихованих шарів, 68 млн. параметрів.

Noisy Speech			
Test set	DS1	DS2	Human
CHiME eval clean	6.30	3.34	3.46
CHiME eval real	67.94	21.79	11.84
CHiME eval sim	80.27	45.05	31.33

Read Speech			
Test set	DS1	DS2	Human
WSJ eval'92	4.94	3.60	5.03
WSJ eval'93	6.94	4.98	8.08
LibriSpeech test-clean	7.89	5.33	5.83
LibriSpeech test-other	21.74	13.25	12.69

Рисунок 2.29 – Залежність між розмірністю датасету та WER

Рисунок 2.29 зображує залежність між розміром датасету та якістю розпізнавання. Дана залежність пояснюється збільшенням репрезентативності вибірки шляхом захоплення більшої кількості

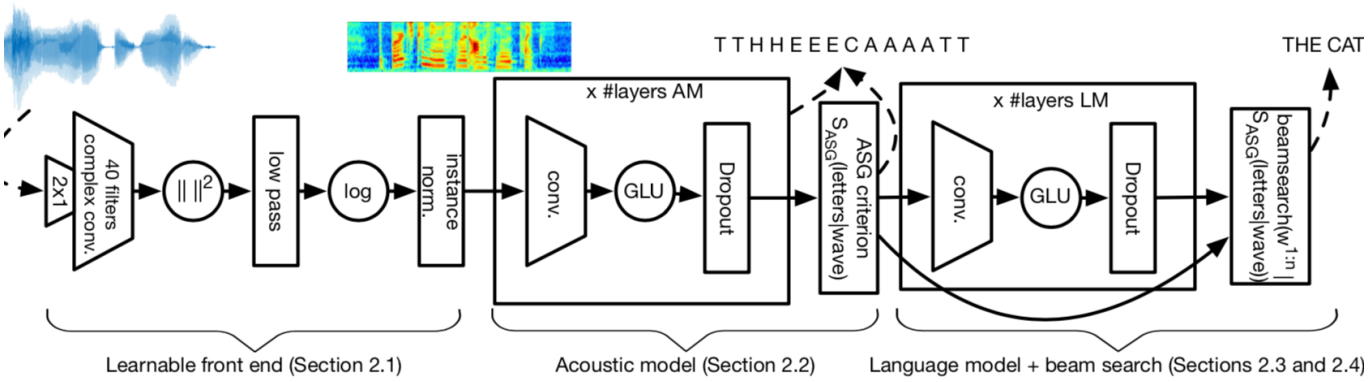
Dataset	Speech Type	Hours
WSJ	read	80
Switchboard	conversational	300
Fisher	conversational	2000
LibriSpeech	read	960
Baidu	read	5000
Baidu	mixed	3600
Total		11940

Fraction of Data	Hours	Regular Dev	Noisy Dev
1%	120	29.23	50.97
10%	1200	13.80	22.99
20%	2400	11.65	20.41
50%	6000	9.51	15.90
100%	12000	8.46	13.59

можливих об’єктів мовлення. Отже, з’ясовано що зі збільшенням розмірності датасету зменшується WER.

Рисунок 2.30 – Порівняння технологій розпізнавання

Рисунок 2.30 надає порівняльні дані якості розпізнавання



комп’ютерних систем. Показовим є факт переваги комп’ютерних систем над людиною у розпізнаванні чіткого мовлення. З іншого боку, люди більш сумісні із розпізнаванням мовлення в умовах наявності завад.

Рисунок 2.31 – Згорткова система голосового розпізнавання

Дана система може бути побудована за допомогою суто згорткової архітектури, що зображено на рис. 2.31.

1. Блок Learnable front end (Section 2.1) – відповідає за трансформацію вхідного сигналу;
2. Блок Acoustic model (Section 2.2) – виконує GLU перетворення, зображене на рис. 2.28-2.29.

3. Блоки Language model та beam search (sections 2.3, 2.4) – виконують роль повністю з’єднаного кінцевого шару.

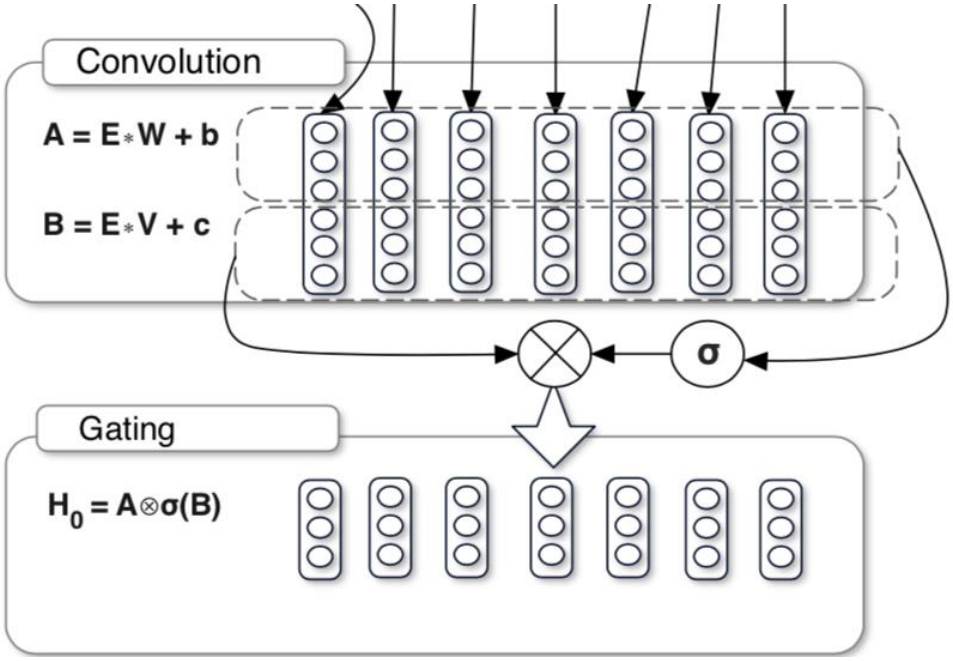


Рисунок 2.32 – Архітектура GLU

2.4 Висновки

В ході проведених досліджень були обрані алгоритми трансформації вхідних даних, перспективні НМ архітектури, сформовано нові вимоги до системи ідентифікації користувача.

Вимоги до підсистем біометричної ідентифікації:

- Здатність до трансформації вхідних даних;
- Наявність НМ архітектури призначеної для вирішення задачі розпізнавання голосового сигналу;
- Здатність НМ до ідентифікації суб’єктів;
- Рівень помилкового розпізнавання менше 5%.

Також, сформовано основну вимогу до кінцевого програмного продукту:

- Наявність механізму ототожнення ідентифікованих суб’єктів з двох підсистем, що реалізується за допомогою бази даних.

3. ОПИС РОЗРОБЛЕНИХ АЛГОРИТМІВ

3.1. Алгоритм візуальної біометричної ідентифікації

Головною проблемою у вирішенні задачі візуальної ідентифікації є верифікація результату. Для її розв'язання застосовуються нейронні мережі, що навчаються на наборах даних задля досягнення відносно безпомилкового рівня розпізнавання.

Введемо поняття target спроби, imposter спроби та об'єктів. Об'єкт $x \in X$, для якого відомо, що $p(A(x) = 1|x) = 1$, тобто об'єкт володіє потрібною властивістю з ймовірністю 1. Об'єкт $x \in X$ для якого $p(A(x) = 1|x) = 0$.

Відповідно, розглянемо підмножини $T = \{x \in X|p(A(x) = 1|x) = 1\}$ та $I = \{x \in X|p(A(x) = 1|x) = 0\}$, які разом формують валідаційну множину $T \cup I = \overline{M}$. Дана множина сформована за міркуваннями репрезентативності – вона відображає необхідну варіативність і має достатній розмір.

Наступним кроком є побудова $F(T) = \{F(t)|t \in T\}$ – результат застосування F до усіх target спроб. Отримана множина дійсних чисел з інтервалу [0,1] репрезентативна вибірка зображена на рис 3.1.

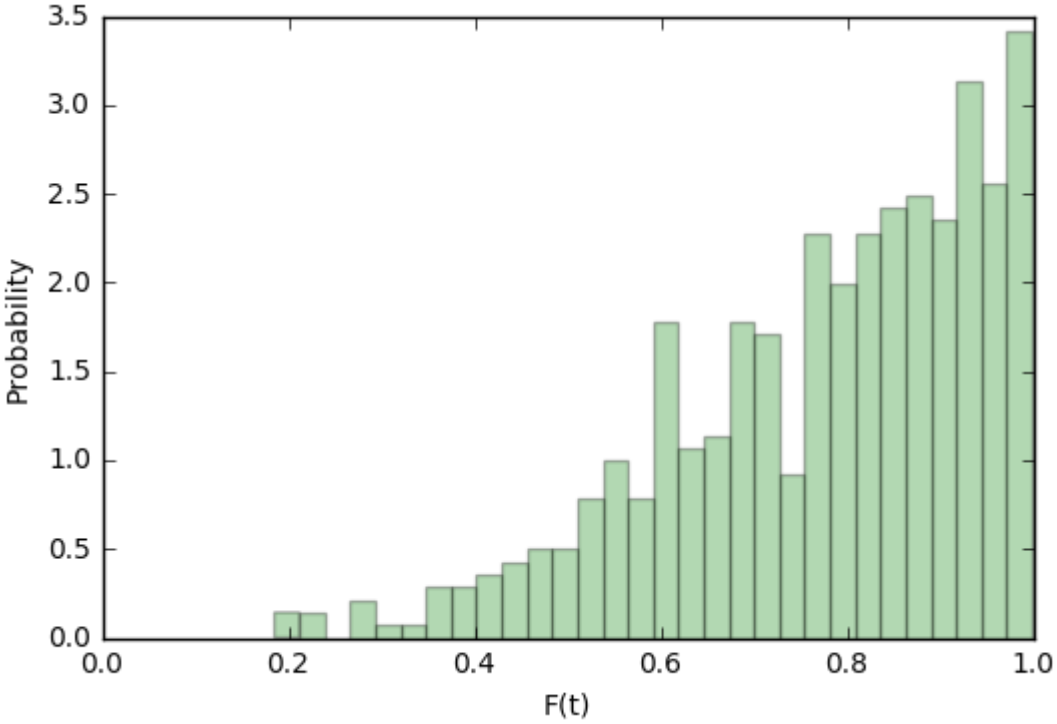


Рисунок 3.1 – Репрезентативна вибірка target спроб

Побудуємо F(I) для оцінки розподілу imposter спроб та подальшого порівняльного аналізу. Рис. 3.2 відображає даний розподіл.

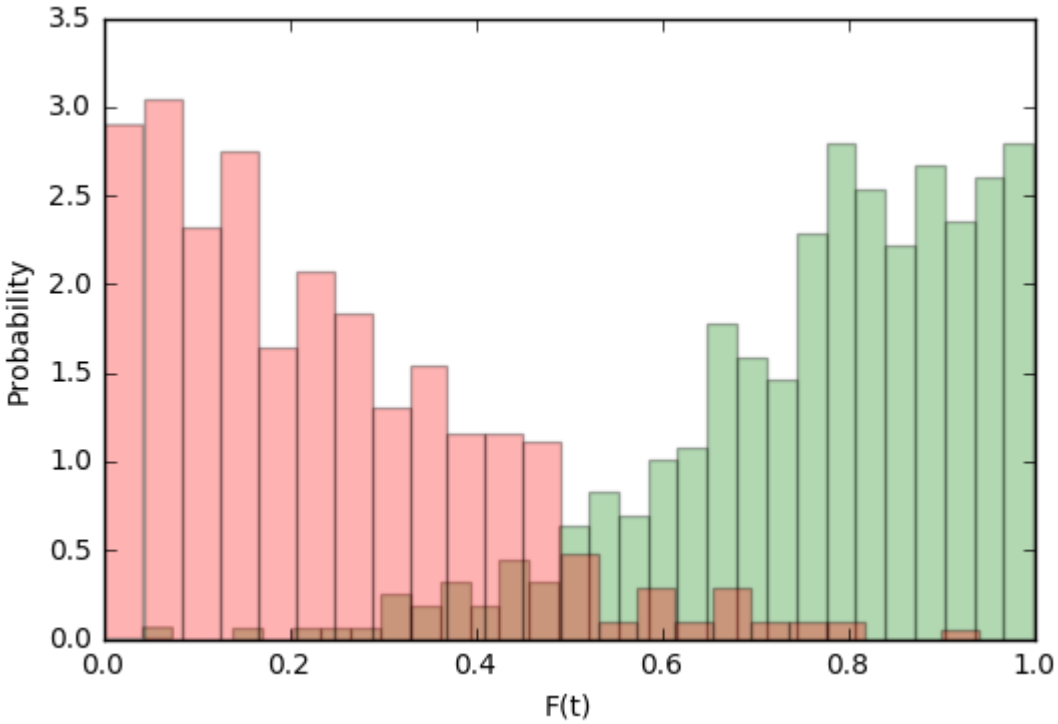


Рисунок 3.2 – Порівняльний розподіл target та imposter спроб

Для надання можливості однозначної ідентифікації застосуємо поняття порогу $d \in [0,1]$, і якщо F(t) < d, результат оцінки вважаємо негативним. $|\{x \in F(T)|x < d\}|$ - кількість target спроб, що будуть невірно класифіковані як imposter спроби. Аналогічно для imposter спроб. Введемо відносну метрику:

$$FRR = \frac{|\{x \in F(T)|x < d\}|}{|F(T)|} \quad (3.1)$$

$$FAR = \frac{|\{x \in F(I)|x > d\}|}{|F(I)|} \quad (3.2)$$

- FRR (False Rejection Rate) – частка помилково відхилених target спроб
- FAR (False Acceptance Rate) – частка помилково прийнятих imposter спроб

Відобразимо FRR та FAR для N точок з постійним кроком $\Delta d = \frac{1}{N}$ з інтервалу [0;1] (рис 3.3):

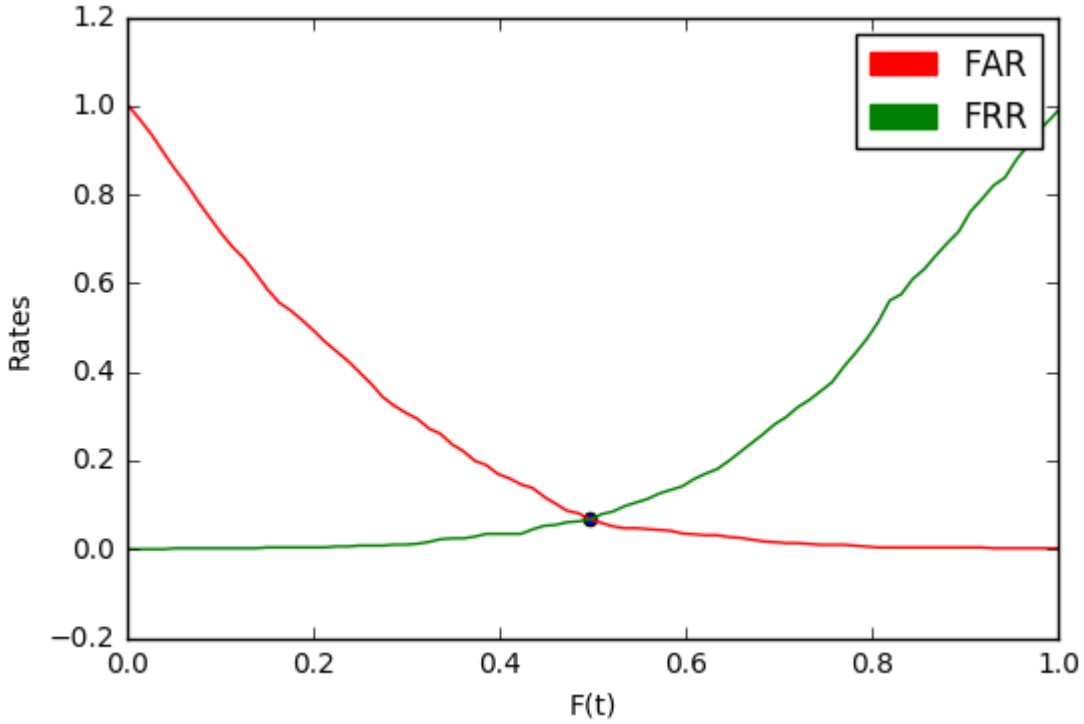


Рисунок 3.3 – Відображення FAR та FRR метрик

В даному випадку точка перетину FAR та FRR є показником EER (Equal Error Rate).

$$EER = \arg\min_{FAR} |FAR - FRR| \tag{3.3}$$

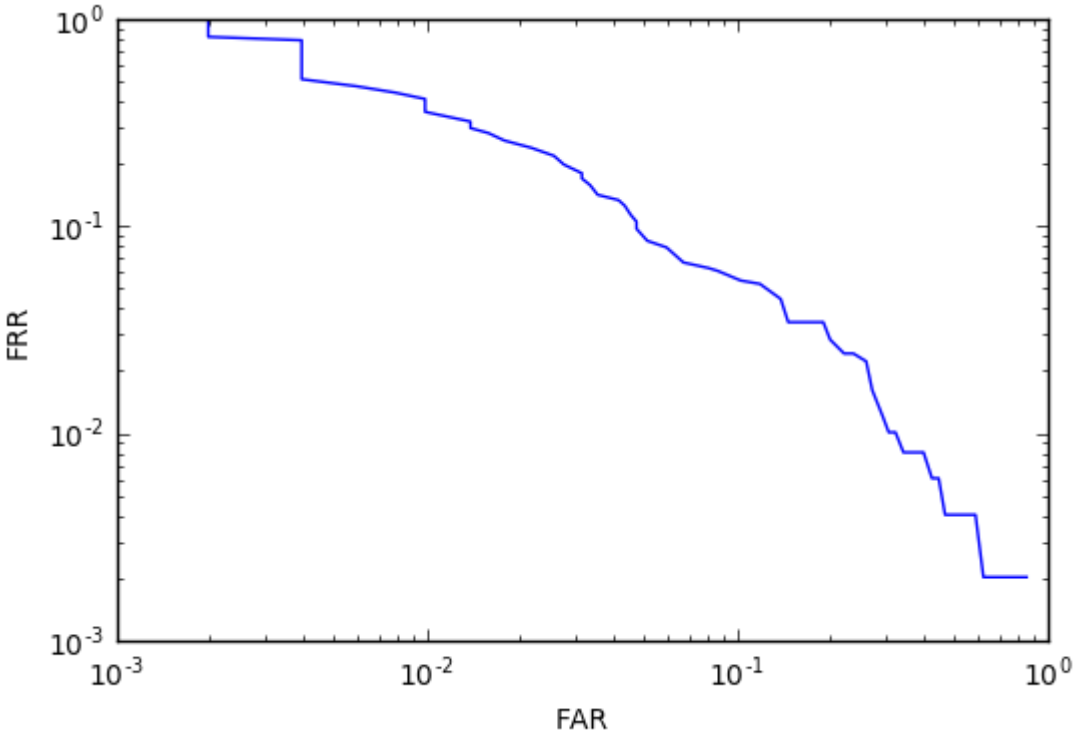


Рисунок 3.4 – DET-крива

EER – об’єктивний критерій оцінки якості ідентифікації, це середня помилка на валідаційній множині. Значення ERR показує частку неправильно прийнятих imposter спроб та невірно відхилених target спроб.

Іншою аналітичною метрикою є DET-крива (рис 3.4), що відображає залежність FAR від FRR в логарифмічному масштабі. За її формою здійснюється ефективна оцінка роботи системи в цілому. EER в даному випадку є точкою перетину DET-кривої з прямою y = x.

Після введення даних метрик, задача контролю вважається вирішеною – незалежно від функції F, наявна можливість побудови та аналізу графіків за показниками FAR, FRR та ERR на валідаційній множині.

Під поняттям валідаційної множини мається на увазі датасет для розпізнавання обличчя. На даний момент наявна велика кількість вибірок з варіативністю по освітленню, віку, положенню обличчя та іншим критеріям.

В рамках виконання дипломної роботи сформовані такі вимоги до датасету:

- доступність в мережі;
- відносно невеликий обсяг;
- наявність варіативності по положенню обличчя та віку суб’єкта.

З проаналізованих в мережі датасетів, критерії задовольнили 3 набори, що були об’єднані в один:

- Caltec Faces;
- FEI Face Database;
- Georgia Tech face database.

В отриманій базі наявно 277 суб’єктів, 4000 зображень, по 14 на особу в середньому. 5-10% суб’єктів взято для development-множини, інші використовуватимуться для навчання, в процесі якого система бачить

лише приклади з другої множини, а її перевірка, шляхом підрахунку EER, відбувається на першій множині.

Рисунок 3.5 – вхідне зображення

Вхідне зображення (рис. 3.5) необхідно опрацювати. Спершу виділимо обличчя.

Інструментарій що застосований при виконанні роботи:

- Мова програмування Python 3.7.3;
- Програмна бібліотека для машинного навчання Tensorflow;

Використовується для цілей побудови і тренування НМ, з метою пошуку та класифікації образів, а також для задач візуалізації, за допомогою фреймворку TensorBoard.

- Прикладний інтерфейс Keras API;

Застосовується для взаємодії з бібліотекою Tensorflow, та розгортання інфраструктури.

- Кросплатформенна бібліотека dlib.

Зберігає в собі інструменти комплексного машинного навчання.

За допомогою dlib отримуємо прямокутник (рис.2.6), що обмежує обличчя суб’єкта з вхідного зображення. Формальна постановка задачі вимагає приведення всіх обличь до одного розміру. Задовольнимо цю вимогу шляхом нормалізації обличь за ключовими точками (очі, ніс, губи).

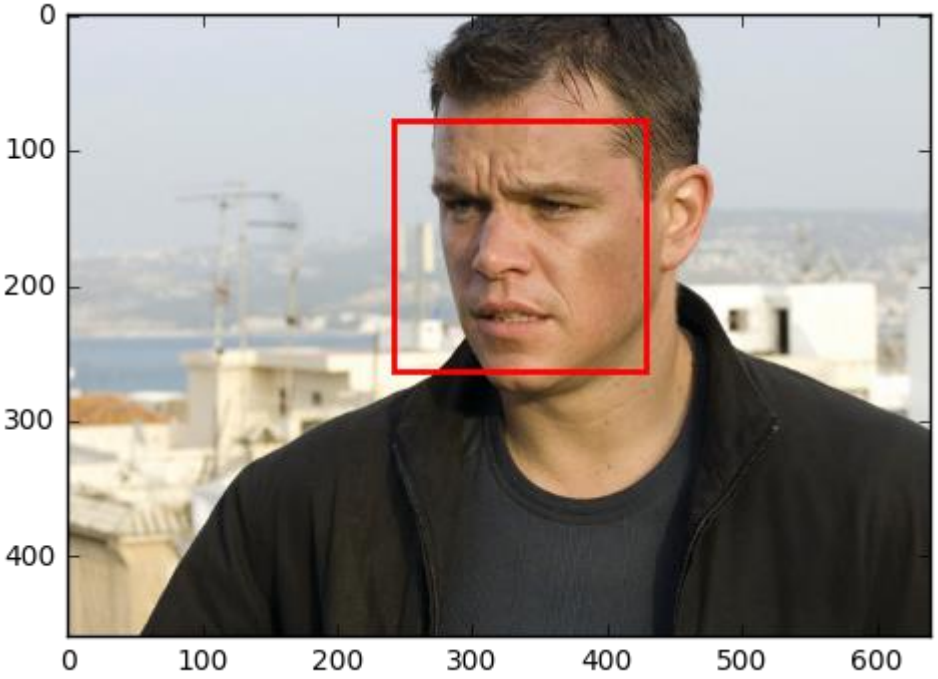


Рисунок 3.6 – Перетворення вхідного зображення

Алгоритм перетворення:

1. Наявні апріорні положення ключових точок в одиничному квадраті;
2. Знаючи обраний розмір зображення, розрахуємо координати ключових точок на зображенні шляхом масштабування;
3. Виділимо ключові точки наступного обличчя;
4. Побудуємо афінне перетворення, що переводить другий набір точок у перший;
5. Застосуємо афінне перетворення до зображення і обріжемо його.

Еталонне положення ключових точок наявне в прикладах dlib (face_template.npy), скористаємося бібліотекою для пошуку ключових точок на зображенні, за допомогою існуючої моделі shape_predictor_68_face_landmarks.dat.

Афінне перетворення однозначно задається трьома точками (рис. 3.7).

Нехай (x_1^0, y_1^0), (x_2^0, y_2^0), (x_3^0, y_3^0) – вхідні точки, які мають бути переведені у (x_1^1, y_1^1), (x_2^1, y_2^1), (x_3^1, y_3^1). Тоді афінне перетворення виражене матрицею T можна знайти із співвідношення.

$$\begin{bmatrix} x_1^1 & x_2^1 & x_3^1 \\ y_1^1 & y_2^1 & y_3^1 \\ 1 & 1 & 1 \end{bmatrix} = T \begin{bmatrix} x_1^0 & x_2^0 & x_3^0 \\ y_1^0 & y_2^0 & y_3^0 \\ 1 & 1 & 1 \end{bmatrix}$$

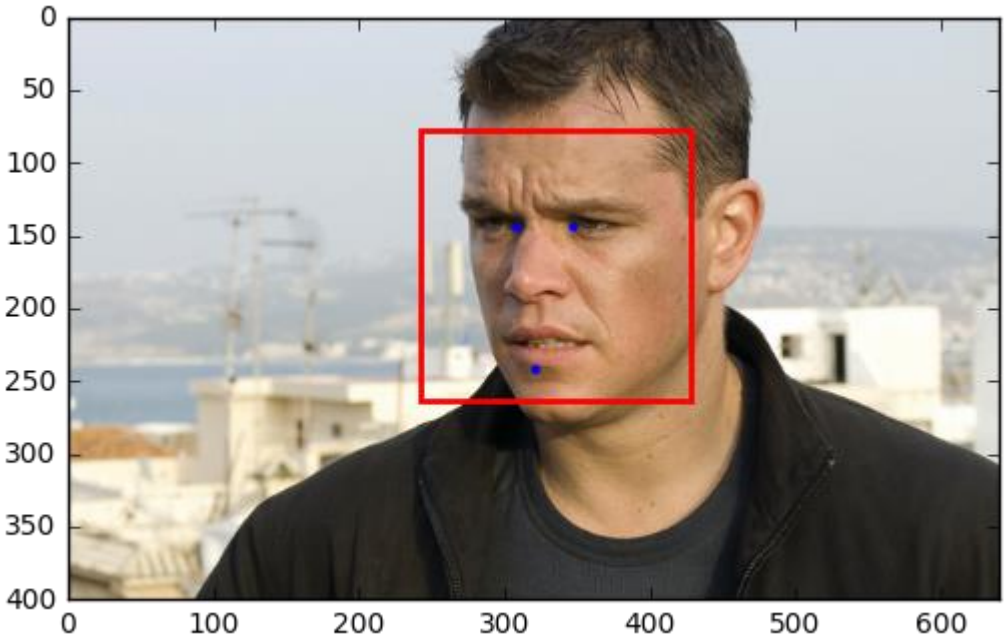


Рисунок 3.7 – ключові точки

Знайдемо T і застосуємо його до зображення за допомогою бібліотеки scipy-image. Фінальним кроком препроцесингу є нормалізація

зображення, шляхом підрахунку середнього та стандартного відхилення по базі навчання і нормування кожного зображення на них (рис.3.8, 3.9).

Рисунок 3.8 – нормування на середньому відхиленні

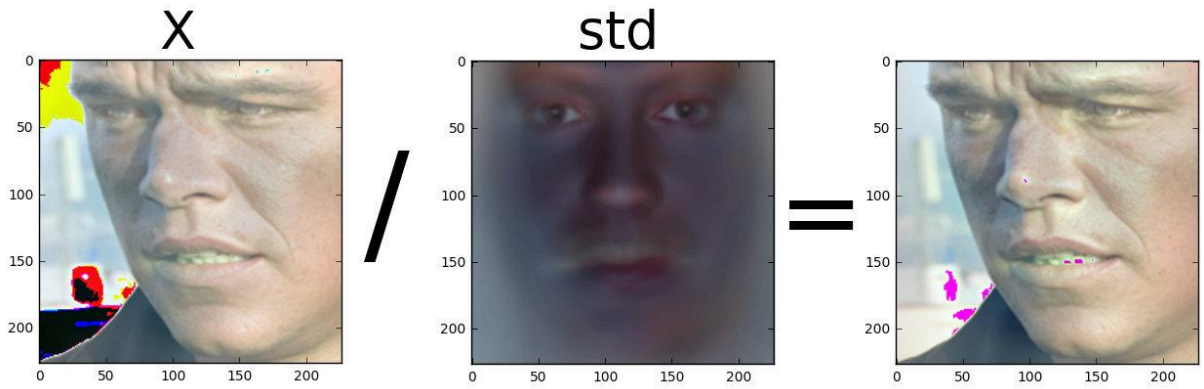


Рисунок 3.9 – нормування на стандартному відхиленні

В ході реалізації системи визначений задовільний рівень EER=10%. Наступним кроком є підбір функції. Застосуємо функцію $F(x, y) = \varepsilon \sim U[0; 1]$. Для кожної пари фотографій датасету отримаємо випадкове значення, побудуємо по ним DET-криву і знайдемо EER (рис. 3.10).

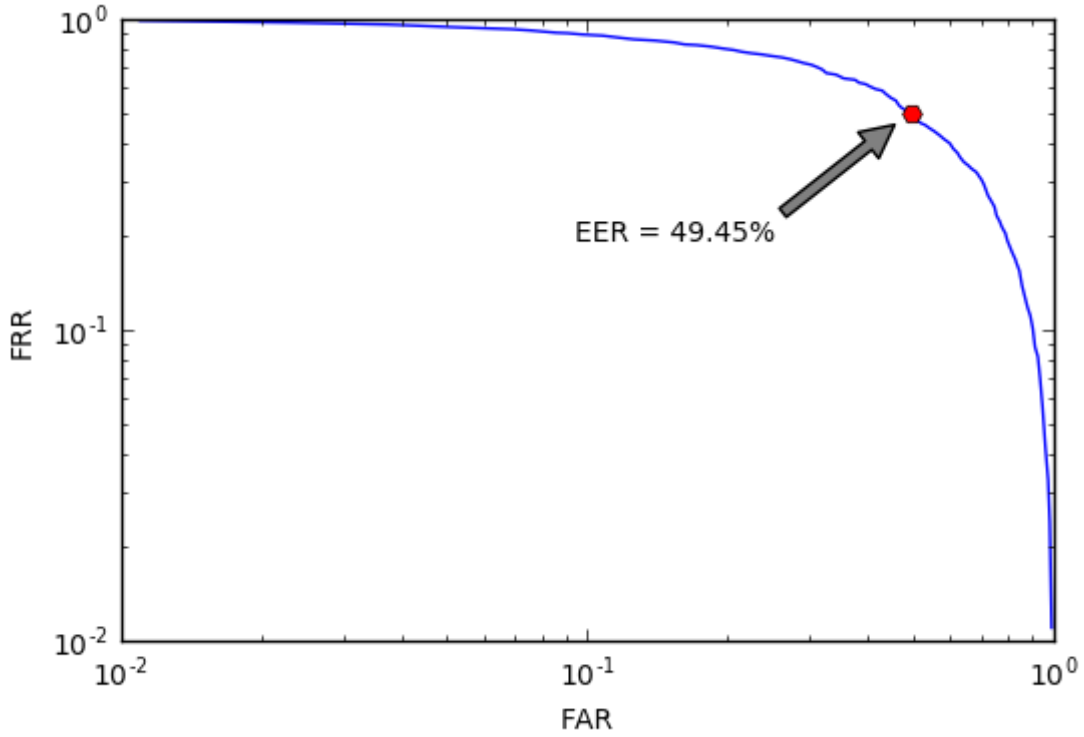


Рисунок 3.10 – DET-крива функції $F(x, y) = \varepsilon \sim U[0; 1]$

В даному випадку EER=49.5%.

Розглянемо метричний простір R^{m*n}. Для будь яких двох його елементів x та y визначено відстань $d(x,y) \in R$, яку можна вводити різними способами. Але розглядати цю відстань слід у від'ємній площині, так як відстань змінюється від нуля до плюс нескінченності, а в прийнятій формалізації має бути навпаки.

Застосуємо косинусну дистанцію до попередніх операцій (рис. 3.11):

$$-d(x,y) = \cos(x,y) = \frac{x*y}{||x||*||y||} \quad (3.4)$$

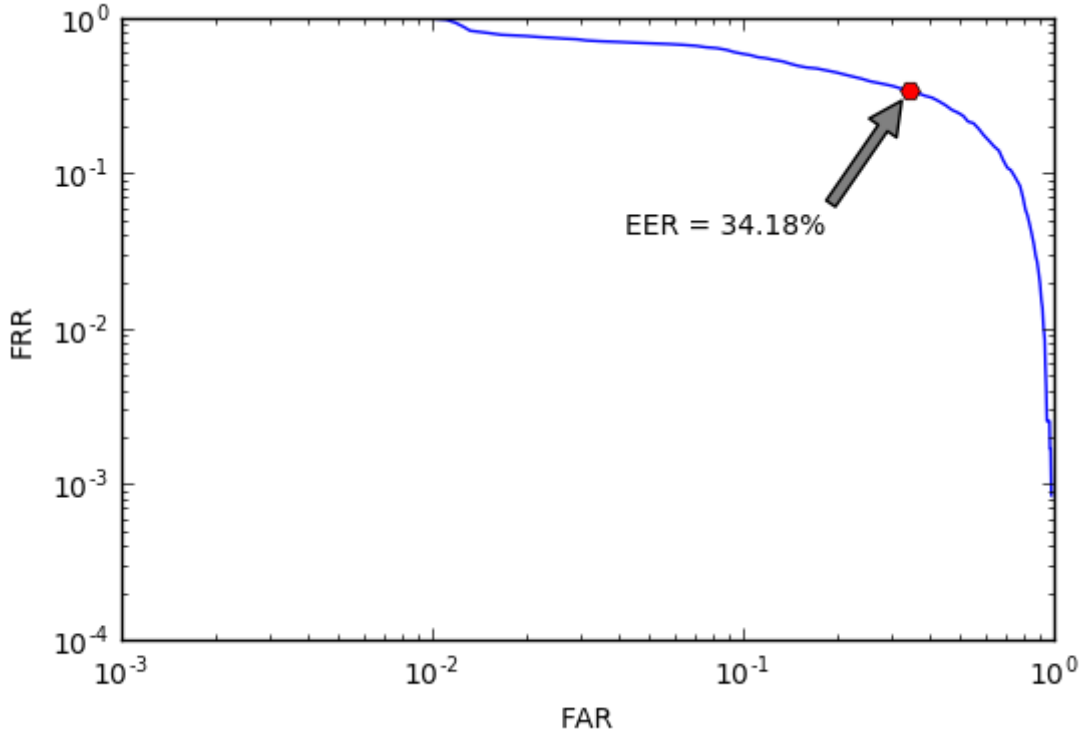


Рисунок 3.11 – DET-крива із косинусною дистанцією

EER зменшився на 16% і став дорівнювати 34.18%. Застосування косинусної дистанції довело перспективність даного рішення, тож подання функції з урахуванням косинусної дистанції:

$$F = d\big(f(x)\big), f(y)) \quad (3.5)$$

- d – косинусна дистанція;

- $f: R^{m*n} \rightarrow R^k$ – функція, що має назву embedder, а результат її роботи з простору R^k – embeddings. Вона вбудовує вхідні зображення в деякий простір іншої розмірності, враховуючи апріорний досвід, добутий із навчальної множини.

Останнім кроком є пошук f, та інтеграція нейромережевої архітектури. На даний момент найпродуктивнішими в питаннях роботи із зображеннями є архітектури згорткових НМ – CNN (Convolutional Neural Networks). Обрана архітектура (рис. 3.12).

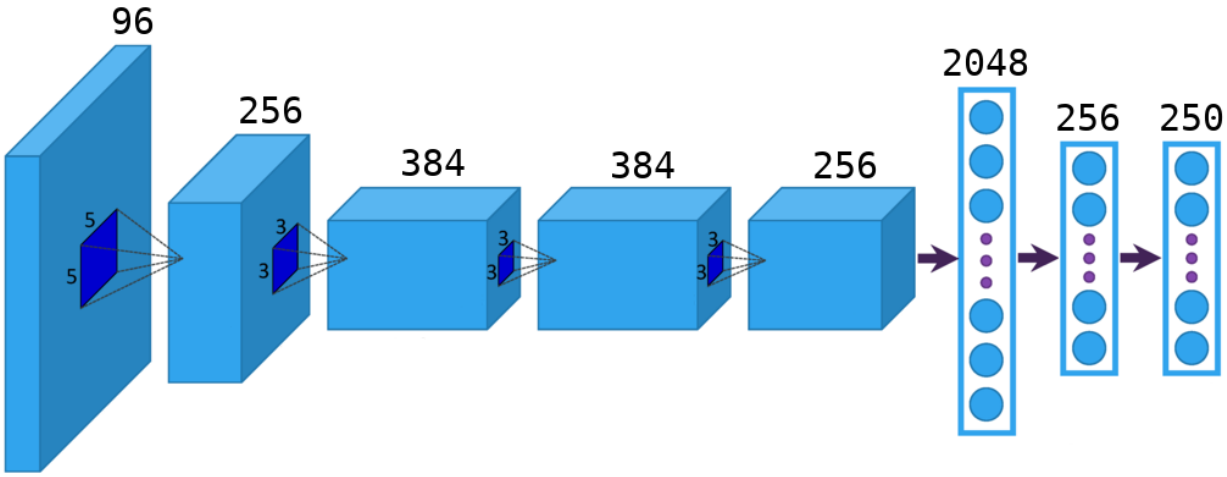


Рисунок 3.12 – архітектура CNN

Навчимо дану модель вирішувати задачу класифікації на навчальній множині: визначати, кому із 250 суб'єктів належить фотографія обличчя. Спершу необхідно застосувати процедуру аугментації, інакше даних не вистачить для задовольняючих результатів. Така техніка вилучення характеристик низької розмірності з останніх слоїв навченої НМ носить назву bottleneck. Після внесених змін аналізуємо DET-криву (рис. 3.13).

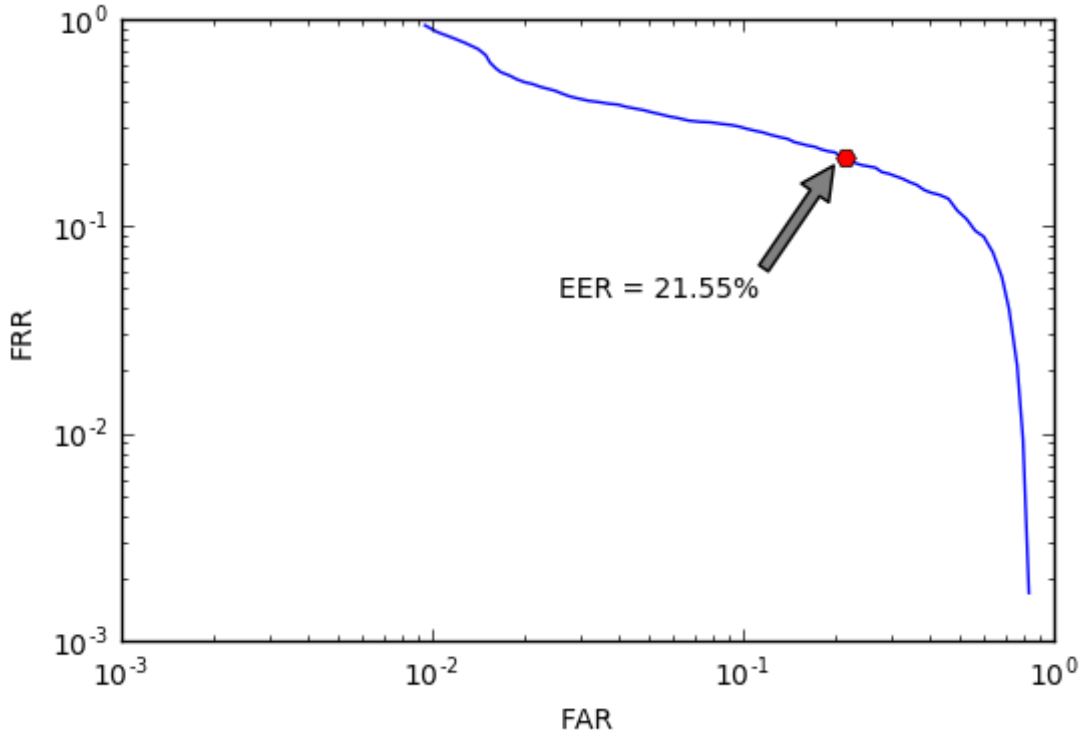


Рисунок 3.13 – DET-крива після інтеграції НМ

Показник EER впав ще на 13%, досягнувши результату 21,6%. Для покращення результату розпізнавання слід:

- зібрати більш повну та варіативну базу;
- збільшити глибину CNN;
- застосувати різноманітні методи регуляризації.

Ключ до покращення результатів – оптимізація f інформацією не лише з навчальної вибірки, а й інформацією з функції d. Слід зафіксувати d і навчати f виходячи з апріорного знання, як embedding-и будуть використані для отримання score (рис. 3.14). Вперше такий підхід було застосовано корпорацією Google в роботі FaceNet: A unified Embedding for

Face Recognition and Clustering, він має назву TDE (Triplet Distance Embedding). Сенс підходу – у побудові f як мережі вхідного простору R^{m*n} в простір embedding-ів R^k, без необхідності вирішення проміжного завдання класифікації. Зафіксуємо d як евклідову відстань і врахуємо його в loss-функції таким чином, щоб вектори одного суб'єкта знаходилися у цільовому просторі якнайближче один до одного, і якнайдалі від векторів інших суб'єктів (рис. 3.15)

Рисунок 3.14 – схематичне зображення архітектури TDE

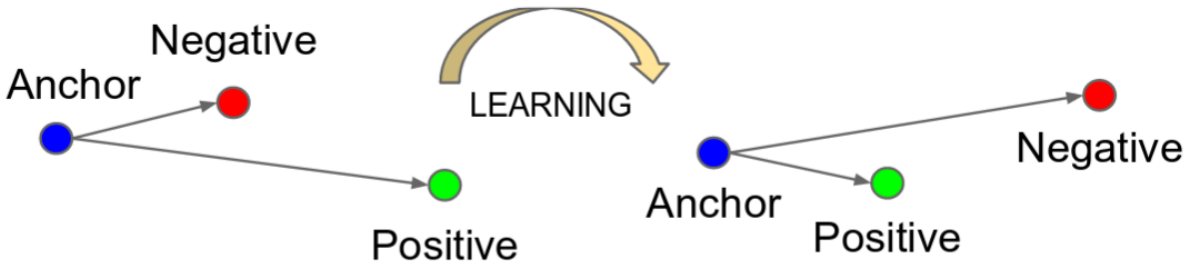


Рисунок 3.15 – методологія TDE

Для навчання поданої мережі застосовуються трійки $(x_a x_p x_n)$, де:

- x_a (anchor);
- x_p (positive);
- x_n (negative).

x_a і x_p – належать одному суб'єкту, а x_n – іншому. Для усіх трьох векторів побудуємо embedding-и $f(x_a)$, $f(x_p)$, $f(x_n)$. Введемо параметр α, і вважаємо що трійка вважається задовільною коли виконується співвідношення:

$$||f(x_a) - f(x_n)||_2^2 - ||f(x_a) - f(x_p)||_2^2 > \alpha \qquad (3.6)$$

Це означає, що для даного anchor між сферами, на яких лежать positive та negative існує проміжок α. Якщо таке співвідношення виконується для всіх наявних трійок навчальної вибірки, то дані вважаються ідеально розділеними. А навчання мережі має сенс лише на трійках, для яких дана нерівність є несправедливою. Виходячи з нерівності, побудуємо loss-функцію для мережі f:

$$L(x_a, x_p, x_n, f) = \frac{1}{N}\sum_{i=1}^{N}[||f(x_a^i) - f(x_p^i)||_2^2 - ||f(x_a^i) - f(x_n^i)||_2^2 + \alpha]$$

(3.7)

Використовуючи даний підхід, рівень помилкового розпізнавання знищується на 30%, але метод має недоліки:

- Потребує дуже багато даних;
- Повільно навчається;
- Невизначеність при підборі α;
- В багатьох випадках (в основному, при малому обсягу даних), архітектура проявляє себе гірше за softmax + bottleneck.

Для подолання цих недоліків застосуємо підхід TPE (Triplet Probalistic Embedding), що нівелює задачу підбору показника α:

$$d(f(x_a), f(x_n)) > d(f(x_a), f(x_p)) \quad (3.8)$$

Даний підхід простіший за початковий та простий в застосуванні – потрібно, щоб найближчий negative-приклад лежав далі за найвіддаленіший positive-приклад, при чому, між ними не обов'язково має бути проміжок. Завдяки тому, що мережа не перестає оновлюватися, коли відстань α досягнута, групи embedding-ів можуть бути перенесені в простір R^k ще краще. Розрахуємо ймовірність того, що триплет задовольняє приведену нерівність:

$$p = \frac{e^{d(f(x_a), f(x_p))}}{e^{d(f(x_a), f(x_p))} + e^{d(f(x_a), f(x_n))}} \quad (3.9)$$

Поділимо на $e^{d(f(x_a), f(x_p))}$:

$$P = \frac{1}{1+e^{d(f(x_a),f(x_n))-d(f(x_a),f(x_p))}} = \sigma(d\left(f(x_a), f\left(x_p\right)\right) - d(f(x_a), f(x_n))$$

(3.10)

Максимізуємо логарифм ймовірності, loss-функція приймає вигляд:

$$L\left(x_a, x_p, x_n, f\right) = -\frac{1}{N}\sum_{i=1}^{N} \log \sigma(d\left(f\left(x_a^i\right), f\left(x_p^i\right)\right) - d(f\left(x_a^i\right), f\left(x_n^i\right))$$

(3.11)

В якості функції f використовуємо f(x)=Wx замість CNN, і навчаємо її на вже отриманих bottleneck ознаках. Отримані результати зображені на

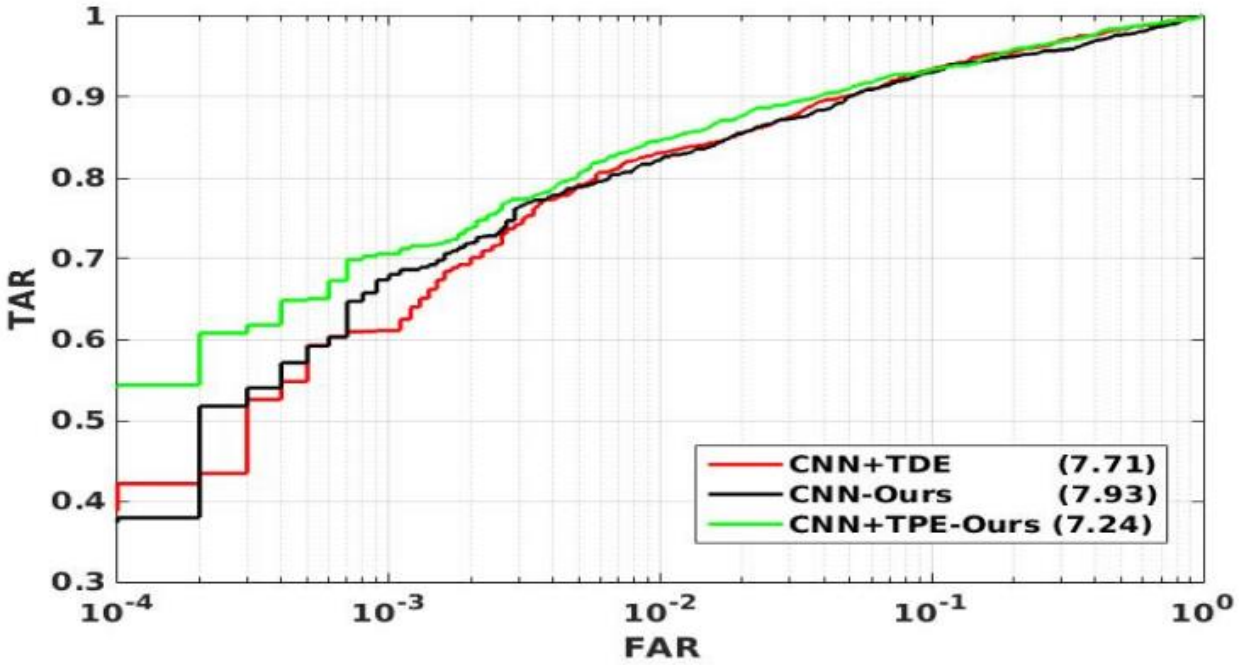


рис. 3.16.

Рисунок 3.16 – порівняльна характеристика застосованих алгоритмів

Даний підхід має такі переваги перед класичним:

- Вимагає невеликої кількості даних;
- Швидко навчається;
- Не вимагає глибокої архітектури;
- Готовий до використання поверх існуючою архітектури.

Навчимо TPE на наявній bottleneck архітектурі та побудуємо аналітичну DET-криву (рис. 3.17). EER = 12%, що вказує на двократну

перевагу перед CNN архітектурами та п’ятикратну перед випадковим вибором.

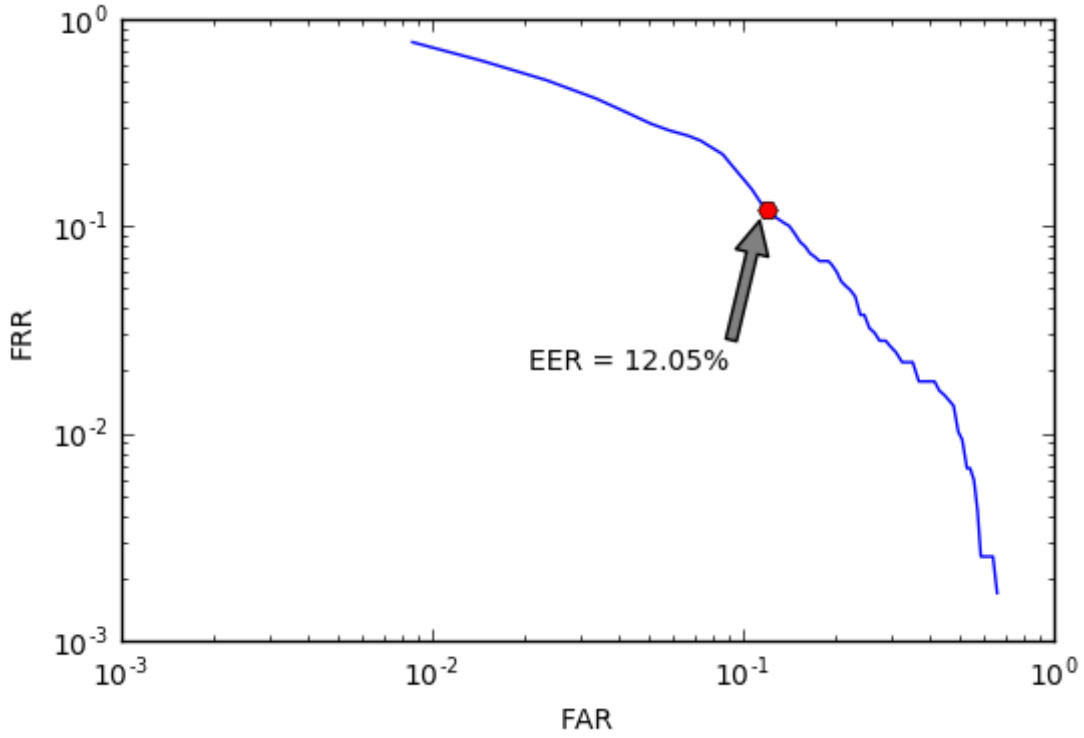


Рисунок 3.17 – фінальний EER

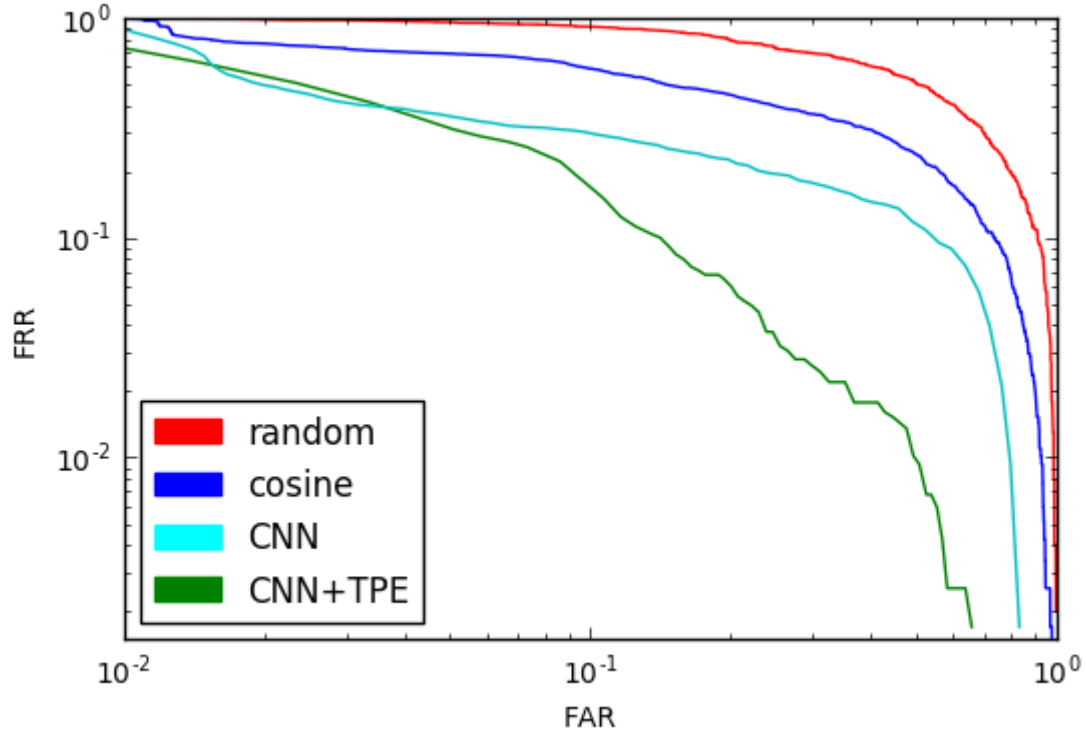


Рисунок 3.18 – порівняльний аналіз усіх розглянутих алгоритмів розпізнавання

3.2. Алгоритм голосової біометричної ідентифікації

В загальноприйнятому розумінні, голос - це динамічна звукова послідовність змінної частоти, що може бути математично інтерпретована як амплітудно-частотна характеристика (рис. 3.19).

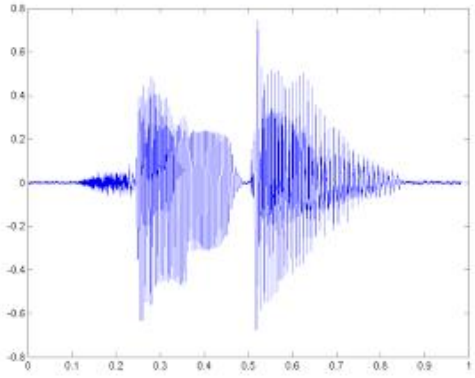


Рисунок 3.19 – Амплітудно-частотна характеристика людського голосу

Людський голос в контексті комп'ютерної системи виражений як спеціальний звуковий тип даних WAV, що є стандартним при роботі з аудіо інформацією (рис. 3.20-3.21).

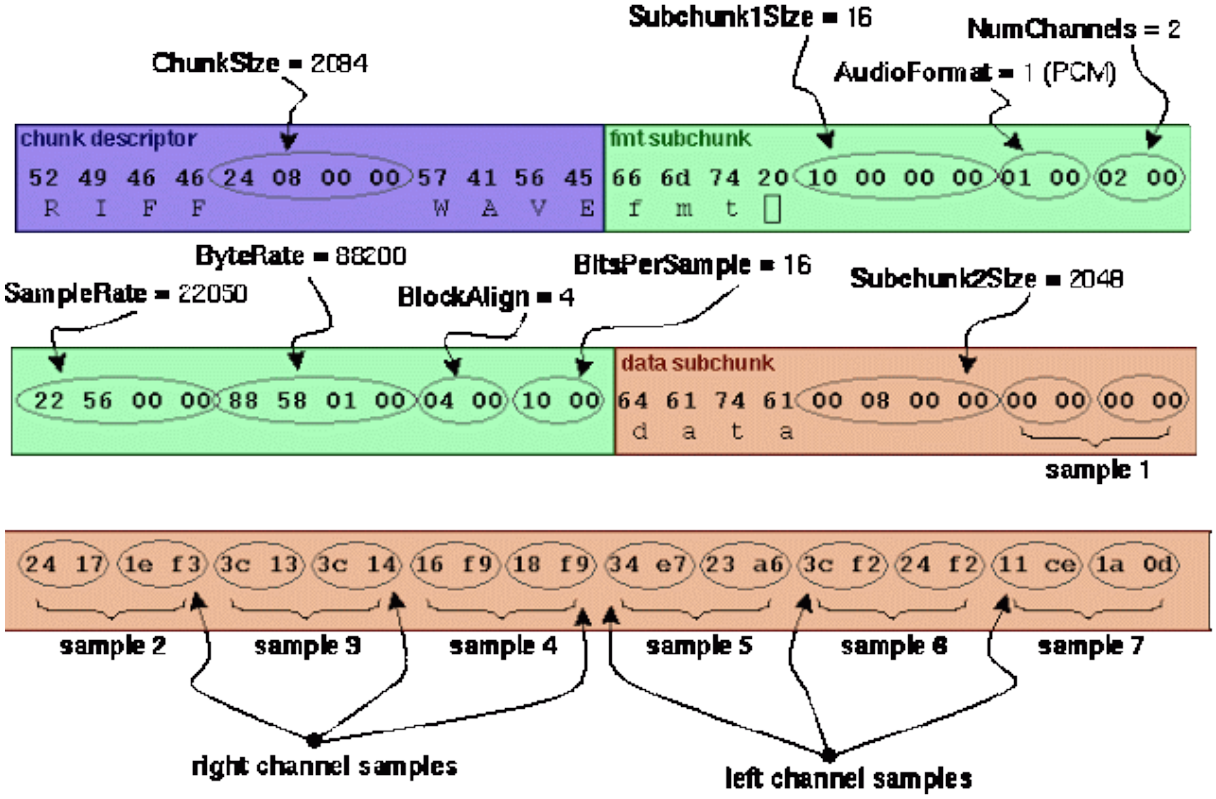


Рисунок 3.20 – Структура формату WAV

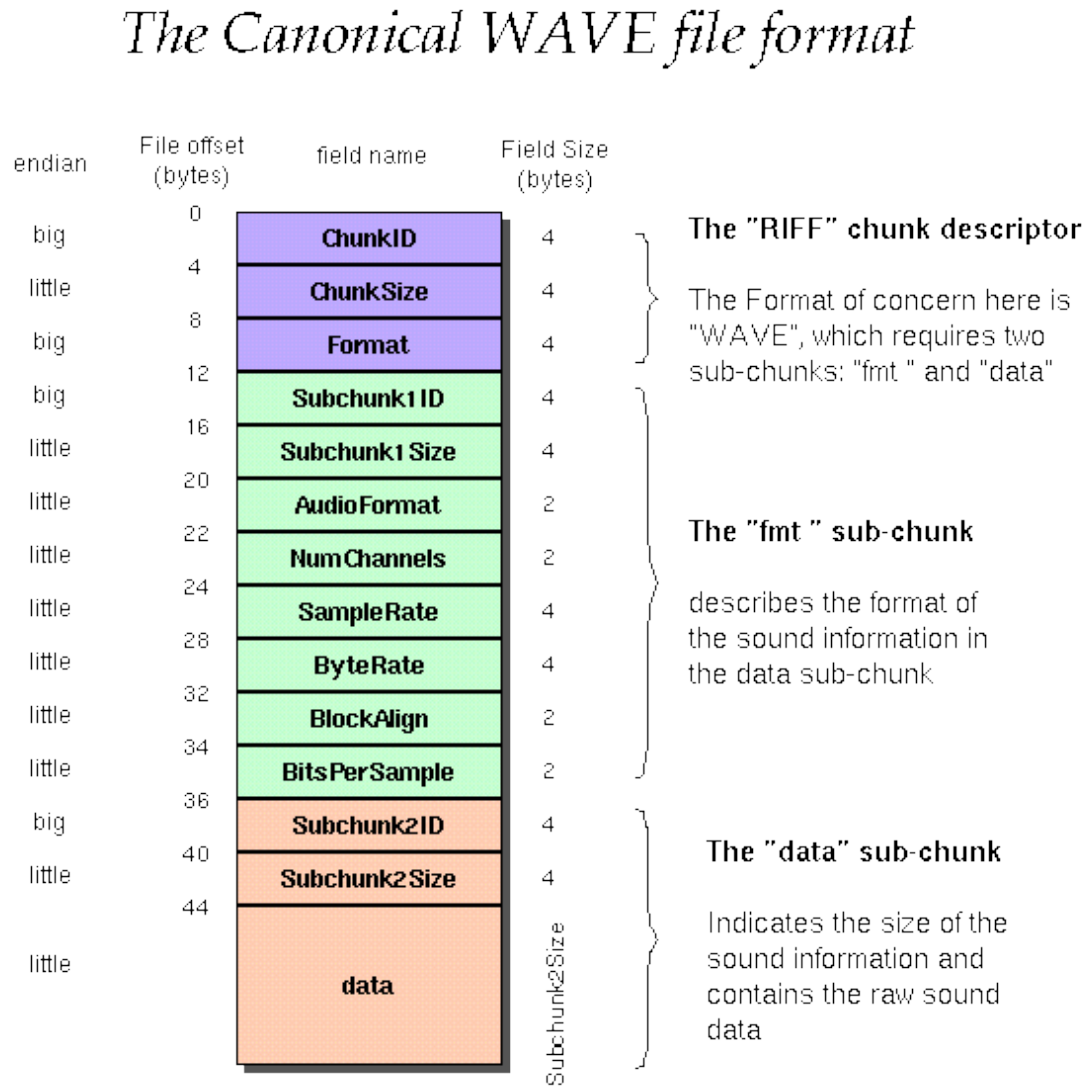


Рисунок 3.21 – Структура WAV формату

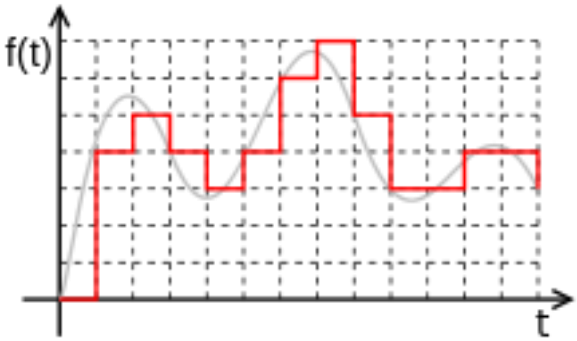
З рисунку 3.21 видно, що даний формат передбачує зберігання даних в двох блоках:

- Першому, що призначений для зберігання інформації про аудіопотік: частота дискретизації файлу, кількість каналів, аудіоформат, бітрейт та інші;
- Другому, що зберігає безпосередньо сам сигнал.

Алгоритм опрацювання вхідних даних:

- Зчитування та аналіз заголовку;
- Збереження даних у виділеному масиві.

Далі, до голосового сигналу застосовується процедура дискретизації, з метою забезпечення подальшого аналітичного процесу (рис. 3.22). Дані поділяються на короткі інтервали часу - фрейми. Фрейми не повинні



слідувати один за одним, вони повинні "перекриватися". Тобто кінець одного кадру повинен накладатися на початок першого.

Рисунок 3.22 – Дискретизація вхідного сигналу

Застосування даної процедури надає можливість аналізу сигналу на інтервалах, замість його дослідження в конкретній точці.

Наступною проблемою є поділ сигналу на слова. Заради спрощення, інтервали мовчання вважатимуться роздільниками слів. В даному випадку, поріг між роздільником і словом задається величиною амплітуди сигналу. Математичне подання порогу – ентропія, що величиною коливання конкретного кадру.

Для розрахунку порогу, в кожному кадрі виконуються наступні дії:

- Нормування сигналу – інтерпретація сигналу в діапазоні [-1;1];
- Побудова гістограми сигналу;
- Розрахунок ентропії за формулою:

$$E = \sum_{i=0}^{N-1} P[i] * log_2(P[i]) \quad (3.12)$$

Задля відокремлення інформативної складової сигналу від тиші, обчислюється поріг ентропії, як середнього між максимальними і мінімальними значеннями всіх кадрів.

Числовою характеристикою обрано RMS – середній квадрат всіх значень.

Наступною складовою алгоритму перетворення сигналу обрано MFCC перетворення, що відображає інтенсивність сигналу.

Переваги даного методу:

- Можливість дослідження частотного спектру сигналу;
- Проекція спектру на мел-шкалу, забезпечує візуальне виділення та підсилення ключових голосових гармоній, необхідних для сприйняття мовлення;
- Можливість регуляції обчислювальних коефіцієнтів, що впливають на кількість досліджуваних гармоній. Внаслідок зменшення кількості коефіцієнтів – відбувається стиснення результуючого сигналу кадру, з метою відсіювання побічних гармоній.

Процес застосування MFCC для фрейму виглядає наступним чином:

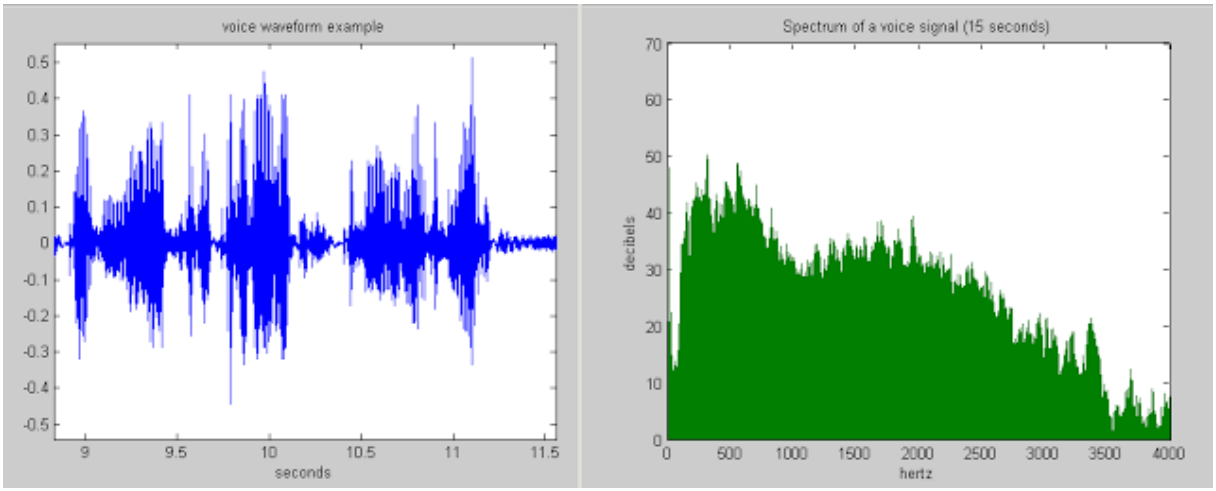
- Фрейм зображується у вигляді вектора $x[k], 0 \leq k < N$, де N – розмір фрейму;
- За допомогою ДПФ відбувається розрахунок спектру сигналу за наступною формулою:

$$X[k] = \sum_{n=0}^{N-1} x[n] * e^{\frac{-2*\pi*i*k*n}{N}}, 0 \leq k < N \quad (3.13)$$

- Застосовується функція Хеммінга задля «згладжування» результатів на краях кадрів:

$$H[k] = 0.54 - 0.46 * \cos(\frac{2*\pi*k}{N-1}) \quad (3.14)$$

- З результуючого вектора що розрахований за формулою 3.15, формується спектрограма вхідного сигналу, яка зображена на рис. 3.23;



$$X[k] = X[k] * H[k], 0 \leq k < N \quad (3.15)$$

Рисунок 3.23 – Перетворення вхідного сигналу у форму АЧХ

- Підрахунок мел-фільтрів. Мел – це психофізична одиниця висоти звуку, що заснована на суб'єктивному сприйнятті середньостатистичного людського індивіда. Значення мел – змінне (рис. 3.24), і залежить від частоти, тембру, гучності звуку. Дана величина застосовується для відображення цінності окремих звуків на спектральному діапазоні. Впровадження мел-трансформації відбувається за формулою 3.16, а зворотне перетворення за формулою 3.17:

$$M = 1127 * \log(1 + \frac{\mathrm{F}}{700}) \quad (3.16)$$

$$F = 700 * (e^{\frac{M}{1127}} - 1) \quad (3.17)$$

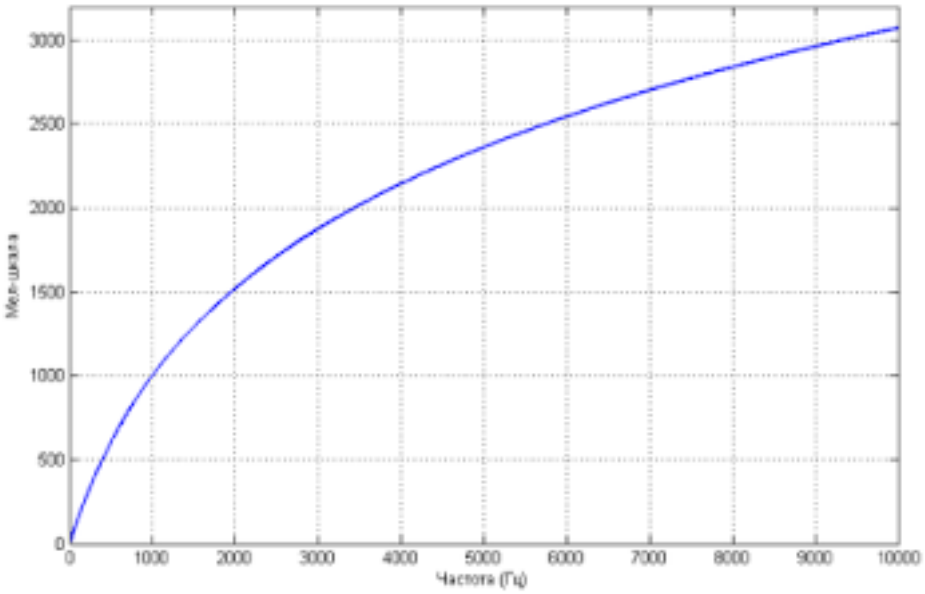


Рисунок 3.24 – Залежність мел-шкали від частоти

При умові наявності сигналу довжиною 256 фреймів, рекомендована кількість мел-коефіцієнтів становить M=10, так як частота дискретизації сигналу має стандарт 16000 Гц, а людська мова лежить в діапазоні [300; 8000] Гц.

Кожне значення мел – це трикутна віконна функція, що підсумовує величину інтенсивності в конкретному частотному діапазоні, формуючи мел-коефіцієнт. Задля розкладення сигналу по всьому частотному діапазону, використовується каскад таких функцій-фільтрів.

Знаючи кількість коефіцієнтів будується набір фільтрів (рис. 3.25):

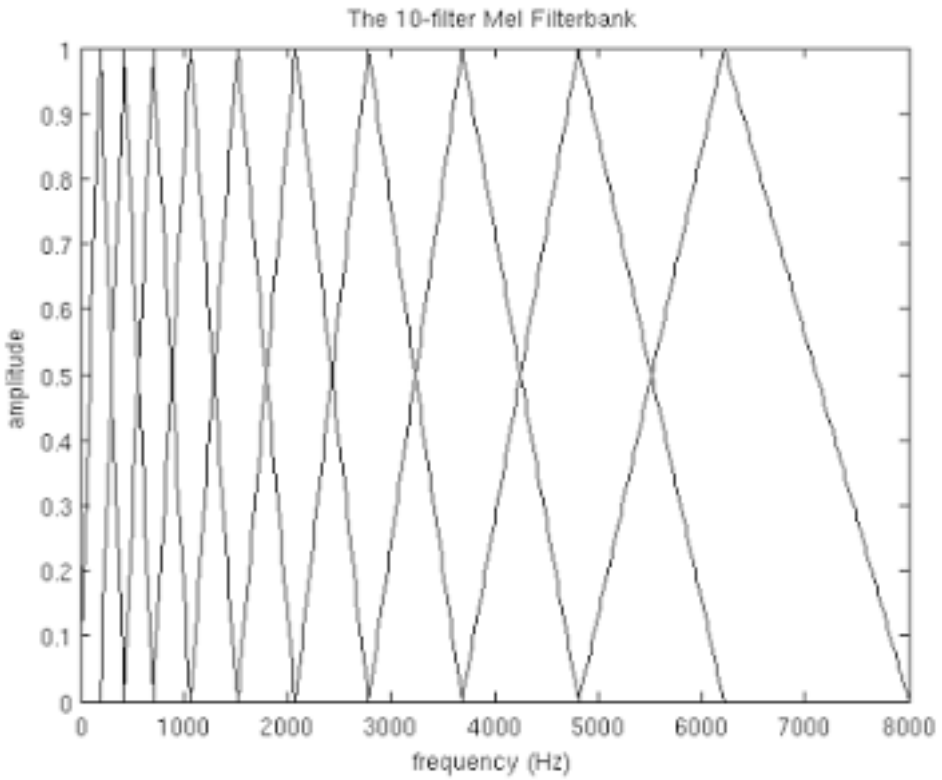


Рисунок 3.25 – Каскад мел-фільтрів

Порядковий номер мел-коефіцієнта визначає ширину базису його фільтра: зі збільшенням серійного номера збільшується ширина базису. Інформативний частотний діапазон знаходиться в межах [300; 8000] Гц. Застосовуючи формулу 3.16, на мел-шкалі даний діапазон перетворюється на [401,25; 2834.99].

Далі відбувається побудова фільтрів за опорними точками: m [i] = [401.25, 621.50, 842.75, 1063.00, 1226.25, 1507.50, 1728.74, 1950.09, 2160.74, 2302.39, 2623.54, 2834.99].

Зворотне перетворення за формулою 3.17 є відображенням даної шкали в частотному діапазоні : h [i] = [300, 516.33, 761.90, 1122.96, 1495.04, 1983.32, 2574.53, 3231.52, 4131.13, 5160.56, 6445.8, 8000].

Як видно з рис. 3.25 збільшення ширини базису фільтра відбувається задля вирівнювання динаміки зростання цінності на різних частотних діапазонах.

Знаючи довжину спектра, його частоту дискретизації, вираховуємо опорні точки за формулою 3.18:

$$f(i) = floor(\ (frameSize + 1) * h(i) \ / \ sampleRate) \tag{3.18}$$

Визначено набір опорних точок:

$$f(i) = 4, 8, 12, 17, 23, 31, 40, 52, 66, 82, 103, 128$$

За даним набором відбувається побудова необхідних фільтрів за допомогою формули 3.19:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \le k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \tag{3.19}$$

Застосування фільтру означає отримання мел-коефіцієнта, і зводиться до попарного множення його значень зі значеннями діапазону, при чому кількість коефіцієнтів дорівнює кількості фільтрів M:

$$S[m] = \log(\textstyle\sum_{k=0}^{N-1} |X[k]|^2 * H_m[k]), 0 \le m < M \tag{3.20}$$

В даному випадку фільтри застосовуються не до самого сигналу безпосередньо, а до значення його інтенсивності. Логарифмування, в свою чергу, зменшує чутливість мел-коефіцієнтів до шуму, зменшує його цінність.

Значення дискретного косинусного перетворення (формула 3.21) полягає у стисканні отриманих результатів, що забезпечує аугментацію інформаційної цінності початкових коефіцієнтів і зменшує цінність кінцевих.

$$C[l] = \textstyle\sum_{m=0}^{M-1} S[m] * \cos(\pi * l * \frac{m+\frac{1}{2}}{M})), 0 \le l < M \tag{3.21}$$

На даному етапі досягнута відповідність між динамічною послідовністю кадрів, та мел-значенням кожного з них, що дає можливість

НМ аналізувати вхідний сигнал.

Побудова НМ починається з формування датасету, що складається з набору даних задіяних у вирішенні задачі. В нашому випадку датасет складається зі зразків голосу.

Цінність датасету полягає у наступних властивостях:

- Після побудови, датасет може бути використаний безліч разів;
- Датасет відображає мовні дані в реальному середовищі, що спрощує аналіз лексично-граматичної структури мови, шляхом наглядного відображення мовленнєвого процесу в динамічну структуру;
- Широка сфера застосування датасетів, що використовуються у тестуванні пошукових систем та машинних морфологій, систем перекладу а також використовуються у лінгвістичних дослідженнях.

Наявність датасету розширяє можливості аналізу мовного матеріалу, та що важливо, надає можливість автоматизування аналітичного процесу. В процесі формування навчальної вибірки, кількість матеріалу визначає значимість, змістовність отриманих даних , а також, рівень їх надійності. Розмітка є ключовим процесом аналізу датасету, що відрізняє навчальний набір від наявних в інтернеті аудіо колекцій, енциклопедій, бібліотек. Текстова розмітка – це механізм відокремлення тексту для більш зручного аналізу.

Існуючі види розмітки:

- Мета-тестова розмітка. Характеризує текст в цілому: назва, автор, дата створення, обсяг, тощо;
- Структурна розмітка – надає інформацію про будову тексту, що дозволяє відокремлювати слова один від одного, визначати межі фрази, речення;

- Лінгвістична розмітка – зображує характеристики мовної інформації: заперечення, питання, тощо.

Повнота і різноманітність розмітки визначає її наукову та освітню цінність.

В Україні навчальна вибірка української мови була сформована співробітниками Інституту філології Київського національного університету імені Тараса Шевченка під керівництвом Н. П. Дарчук.

Датасет складається з текстів, що пропущені крізь лінгвістичний аналізатор, що присвоює кожній структурній одиниці тексту супутню інформацію: сенс, лексичне значення, синтаксична функція в реченні, граматична форма, тощо.

Навчальна вибірка провадить наступні види інформації:

- Конкорданс – контекст використання структурної одиниці. Конкорданс надає можливість дослідження індивідуального авторського використання словосполучень, особливостей використання слів в умовах різного стилістичного забарвлення тексту. Використовується для контекстуального розкриття авторського бачення образів та понять;
- Кількісна характеристика використання мовних одиниць.

Частотна інформація відображає взаємозв’язки лексичної та статистичної складових тексту, семантичні ролі структурних одиниць, їх граматичні особливості.

На даному етапі сформовано навчальний набір ключових WAV-файлів, що забезпечені структурною розміткою.

Навчання НМ відбувається за алгоритмом, що зображений на

кресленні Д2 (див. Додаток 1). Йому передує алгоритм трансформації вхідних даних, що зображений на кресленні Д1 (див. Додаток 1).

3.3.Висновки

На даному етапі досліджені алгоритми реалізовані на базі двох рішень:

- Система біометричної ідентифікації на основі індивідуальних візуальних показників користувача;
- Система біометричної ідентифікації на основі індивідуальних голосових показників користувача.

Сформовано нові вимоги:

- Імплементація розроблених рішень в кінцевий програмний продукт з можливістю забезпечення доступу до бази даних сформованих суб’єктів;
- Тестування та оптимізація системи з метою доведення до загального рівня помилкового розпізнавання менше 10%.

4. АНАЛІЗ РОЗРОБЛЕНОЇ СИСТЕМИ

4.1 Характеристики КС

Інтерфейс розробленої комп'ютерної системи реалізовано у вигляді клієнтського додатку, що зображений на рис. 4.1 :

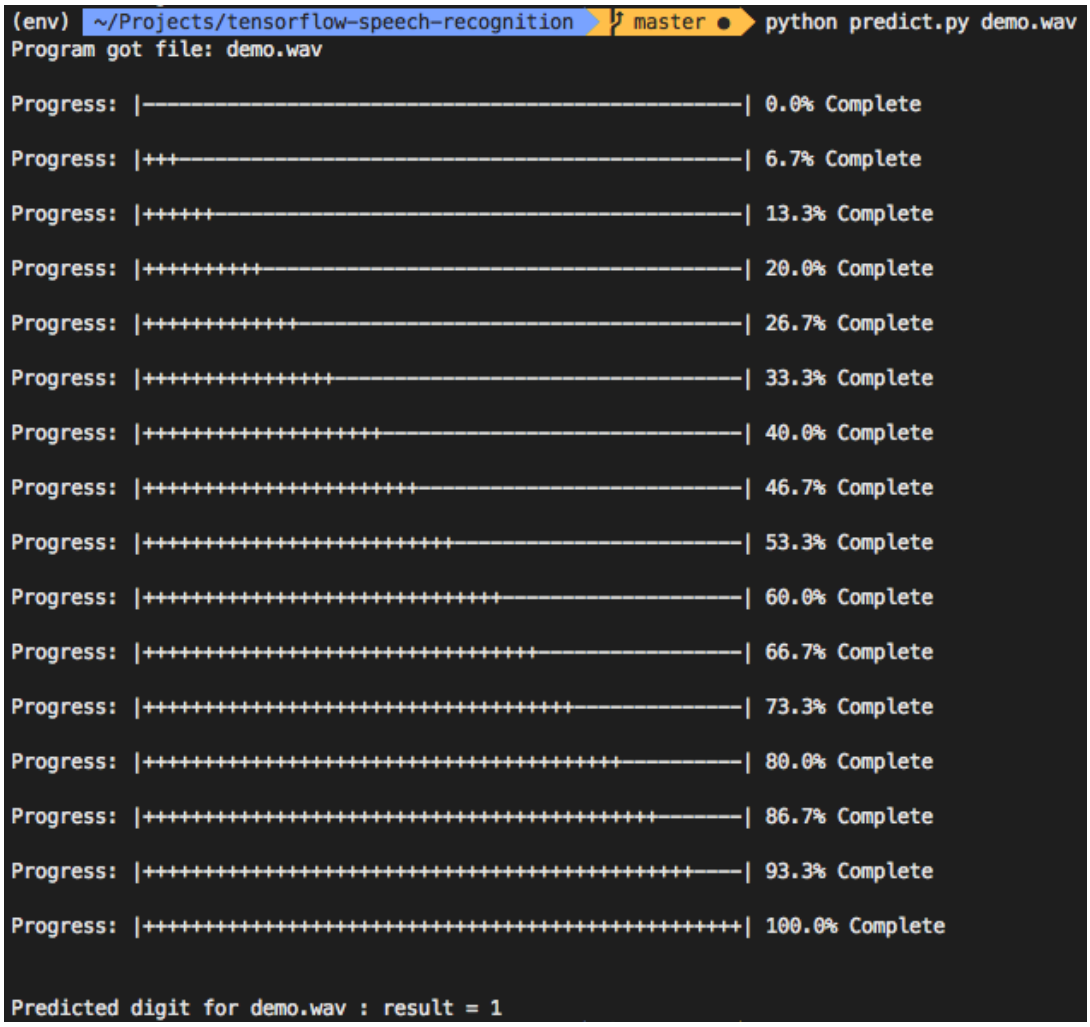


Рисунок 4.1 – Інтерфейс клієнтського додатку

Навчання обох підсистем відбувається наступнім чином. Після виконання команди початку навчання, система автоматично переглядає поточний каталог даних, що зберігає в собі досліджувані дані. При наявності файлів система починає процес навчання. В процесі навчання, програма відображує дані про навчальний прогрес, інформацію щодо рівнів помилкового розпізнавання, час затрачений на кожну ітерацію процесу (рис. 4.2).

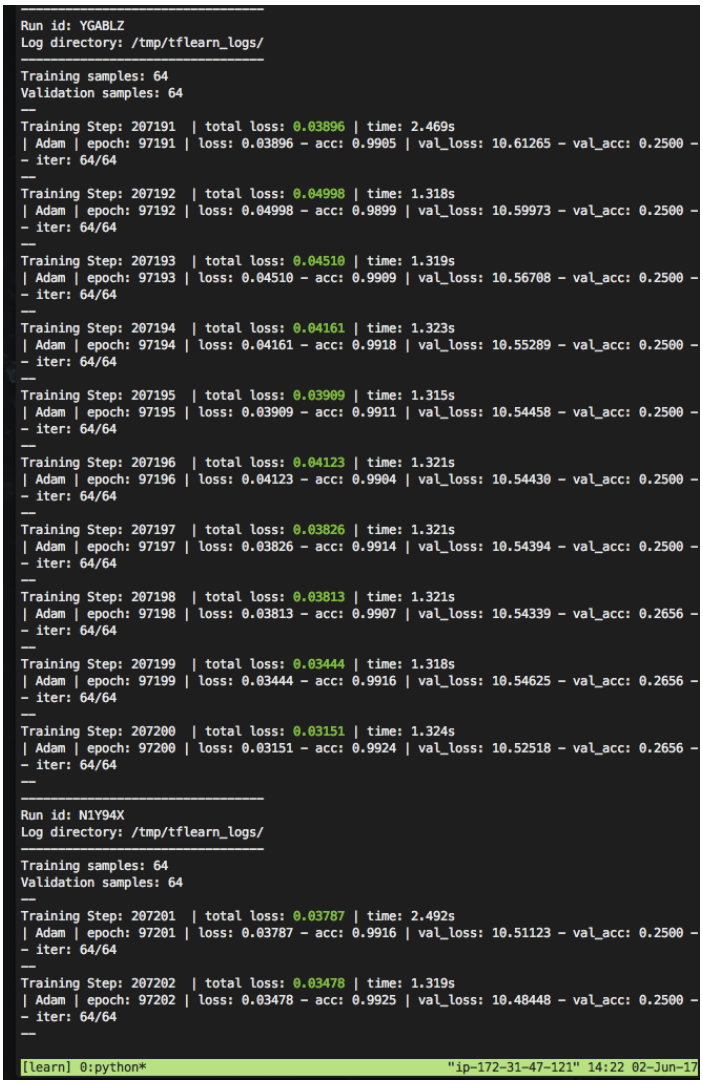


Рисунок 4.2 – Лістинг процесу навчання НМ

Навчальні дані в процесі тренування записуються у визначений додатком Tensorflow каталог /tmp/rflearn_logs. Даний додаток надає можливості дослідження процесу тренування та використання розроблених НМ за допомогою клієнтського інтерфейсу CLI Tensorboard. Даний інструмент відображає інформацію про точність, втрати та інші характеристики процесу навчання та застосування НМ (рис. 4.3).

В додатку наявна можливість перегляду побудованої архітектури у вигляді інтерактивного графу зі згрупованими вершинами (рис 4.4).

Комп’ютерна система розроблена на базі мови програмування Python з використанням бібліотеки машинного навчання Tensorflow, клієнтського

інтерфейсу Tensorboard, та пакетного менеджера PIP.

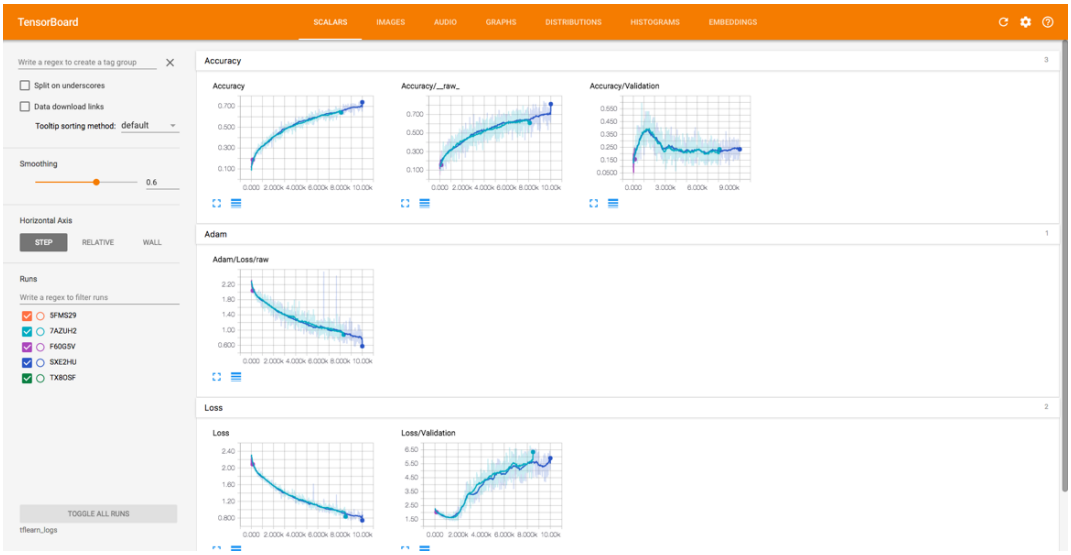


Рисунок 4.3 – Відображення процесу навчання додатком CLI Tensorboard

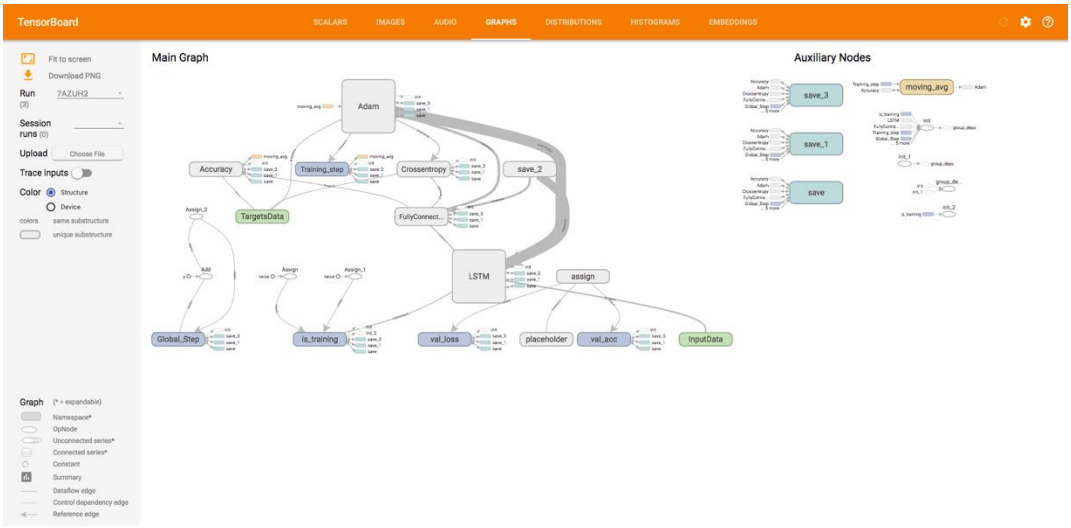


Рисунок 4.4 – Блок-схема розробленої НМ

4.2. Оптимізація КС

Розроблена КС ідентифікації біометричної суб'єкта на основі голосового сигналу, в процесі навчання на наборі даних із 1100 прикладів, та кількістю ітерацій – 10000 показала результати помилкової ідентифікації $8{,}2*10^{-6}$. Даний рівень помилкової ідентифікації задовольняє поставленим умовам.

КС біометричної ідентифікації суб'єкта на основі візуальних показників продемонструвала наступні результати:

Результати тренування КС візуальної ідентифікації суб’єкта Таблиця 4.1

	Загальна кількість обличь	Ідентифіковані обличчя
Вірно ідентифіковані	226	203
Помилково неідентифіковані	226	10
Помилково ідентифіковані	226	13
Точність ідентифікації		89,8%

Отриманий рівень точності ідентифікації менше за 90%, що суперечить поставленим умовам. На даному етапі визначена потреба оптимізації КС шляхом модифікації алгоритму тренування НМ.

Першим підходом до вдосконалення алгоритму є застосування методу регулювання контрастності до вхідних зображень за допомогою рівняння 4.1. Даний метод протестований з різними значеннями альфа та бета, щоб вибрати той, який дає найкращий результат точності виявлення та розпізнавання, тобто значення 1,5 (α) та 0,0 (β).

$$g(x, y) = \alpha * f(x, y) + \beta \quad (4.1)$$

Другим підходом є порівняння ефектів трьох типів фільтрів на точність: Гаусовий фільтр розмивання, медіанний фільтр та двосторонній. Експериментальним шляхом визначено доцільність застосування двостороннього фільтра, що показує найбільшу ефективність в задачах ідентифікації та розпізнавання обличь. Формула 4.2 застосовує

двосторонній фільтр до вхідного зображення.

$$F(x,y) = \frac{\sum_{x=-N}^{N}\sum_{y=-N}^{N} I(x,y)W(x,y)}{\sum_{x=-N}^{N}\sum_{y=-N}^{N} W(x,y)} \quad (4.2)$$

Де:

- W(x,y) – функція зважування фільтра;
- I(x,y) – множина пікселів вхідного зображення;
- F(x,y) – результат застосування двостороннього фільтра до множини 2N+1.

Наступним кроком є застосування CF(x,y) (формула 4.3) з метою зменшення шумів і контролю ефектів контрасту у вхідних зображеннях, де:

- g(x,y) – контрастне зображення;
- F(x,y) – застосований фільтр.

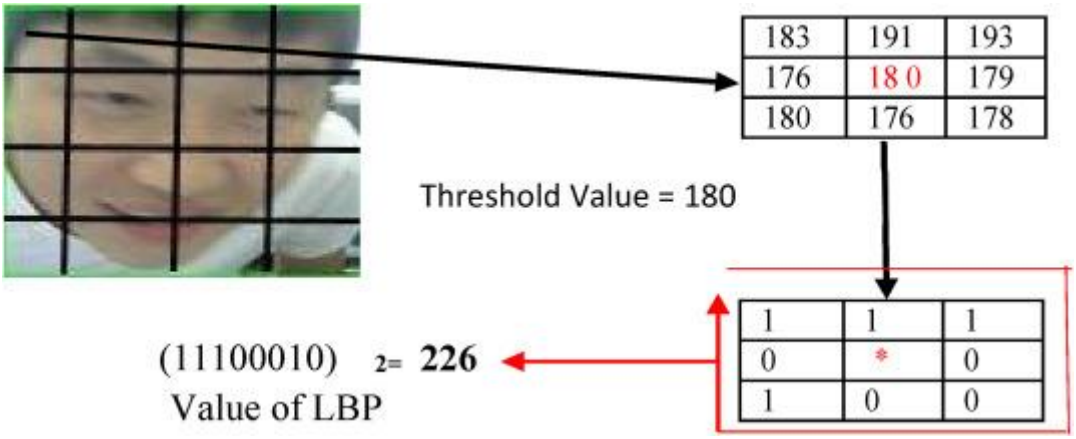
$$CF(x,y) = g(x,y) * F(x,y) \quad (4.3)$$

Результуючі пікселі зображення, що отримані з вищевказаного рівняння фільтруються за допомогою методу гістограмної фільтрації картинки, визначеного в рівнянні 4.4.

$$Eq = H`(CF(x,y)) \quad (4.4)$$

H` - нормований кумулятивний розподіл з максимальним значенням 255.

Тепер алгоритм LBP може бути застосованим до виявлених у зображенні об’єктів з метою їх подальшого вилучення та порівняння. Перший LBP оператор працює з множиною пікселів 3*3, як показано на



рисунку (4.5).

Рисунок 4.5 – Застосування оператора LBP

Більш формальний опис оператора LBP поданий у формулі 4.5.

$$LBPp,r\,(X_c,Y_c) = \sum_{p=0}^{p-1} 2^p S(i_p - i_c) \tag{4.5}$$

Де:

- (X_c, Y_c) – значення центрального пікселя;
- i_p, i_c – значення інтенсивності множини пікселів;
- p – сусідні пікселі в радіусі r навколо досліджуваної множини;
- $S(x)$ – знакозмінна функція визначена формулою 4.6, що застосовується з метою задання порогу.

$$S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4.6}$$

При застосуванні даного підходу були використані формули 4.1 – 4.6, з метою покращення загальної якості вхідних зображень обличчя, та як наслідок, підвищення точності алгоритму ідентифікації LBP (Рис. 4.6).

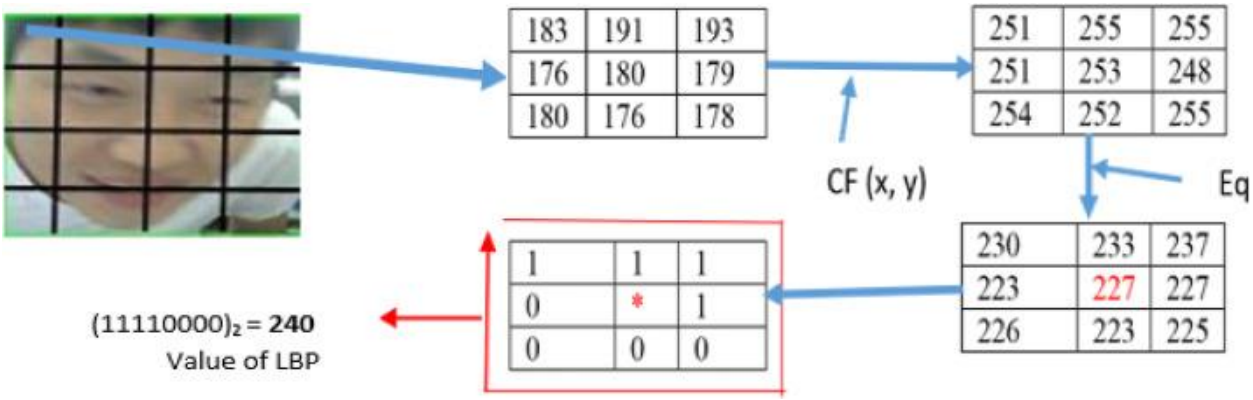


Рисунок 4.6 – Застосування модифікованого оператора LBP

Дослідження кожного вікна у форматі підмножин розмірністю 3*3 пікселі вирішує проблеми наявності шуму, поганої освітленості, різкості, роздільної здатності. Після застосування формул 4.1-4.4, отримані вищі піксельні і порогові значення (рис. 4.6), що надають здатність до більш точного порівняння зображень, та зменшують рівень помилкової ідентифікації.

Таким чином, отримано покращений алгоритм розмітки зображень, що стимулює загальну точність ідентифікації (рис. 4.7).

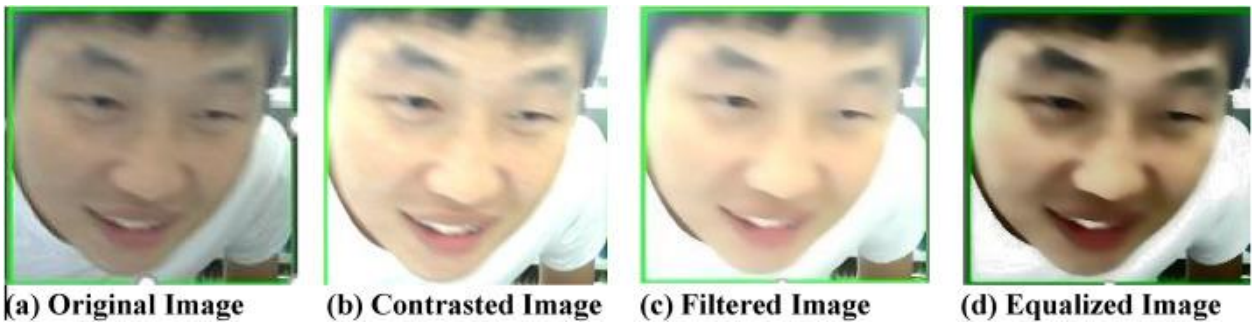
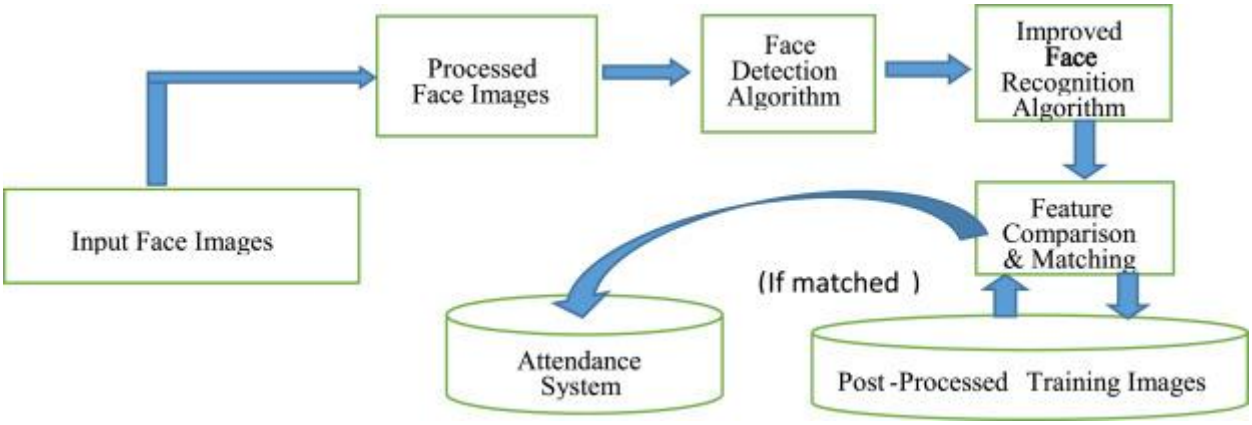


Рисунок 4.7 – Порівняння вхідного і перетворених за допомогою алгоритму LBP зображень

З урахуванням імплементованого алгоритму розмітки LBP, спрощена блок-схема розробленої КС біометричної ідентифікації користувача на основі візуальних показників приймає наступний вигляд



(рис. 4.8):

Рисунок 4.8 – Блок-схема КС біометричної візуальної ідентифікації користувача

Результати тестування оптимізованої КС на різних наборах даних відображено на таблиці 4.2:

Результати тестування оптимізованої КС

Таблиця 4.2

Задіяно обличь	Помилково неідентифіковані	Невідомі обличчя	Помилково ідентифіковані
355	18	1	32
357	6	3	24
363	27	0	37
417	7	4	49
371	10	5	35

Отримана точність розпізнавання 90,49%, що задовольняє

поставленій задачі.

4.3. Висновки

На фінальному етапі розробки КС відбувалося тестування побудованої архітектури, в ході якого була визначена потреба в оптимізації алгоритму візуальної біометричної ідентифікації суб’єкта, заради досягнення поставленого порогу помилкового розпізнавання. Дана проблема була вирішена імплементацією алгоритму LBP, для надання додаткової інформаційної розмітки. Цей факт доводить перспективність методу модифікації математичної моделі для збільшення точності розпізнавання.

Остаточна точність розпізнавання розробленої КС становить 90,49%. Вона обмежена показником точності системи візуальної ідентифікації. Модифікація системи можлива за наступними методиками:

1. Збільшення обсягу вибірки;
2. Збільшення репрезентативності вибірки;
3. Модифікація математичної моделі;
4. Модифікація апаратної бази.

Розроблена КС здатна до ідентифікації суб’єктів на основі біометричного аналізу з точністю 90,49%, що задовольняє поставленим умовам.

ВИСНОВКИ

Дана магістерська дисертація аналізує та вирішує проблему біометричної ідентифікації на основі антропометричних показників суб’єкта, шляхом розробки КС на основі технології НМ. У процесі розробки сформовані наступні результати:

1. Розроблена архітектура для розпізнавання та ідентифікації суб’єктів на основі індивідуальних антропометричних показників:
    - Голосу;
    - Обличчя.
2. Розроблено математичні моделі ідентифікації антропометричних показників на основі технологій НМ;
3. Проведені експериментальні дослідження, що доводять перспективність застосування запропонованої технології біометричної ідентифікації суб’єктів;
4. Розроблена КС рекомендована до застосування в корпоративних умовах з метою ведення обліку ненадійних суб’єктів, та запобігання викраденню даних;
5. Доведено, що перспективним способом оптимізації точності розпізнавання КС на базі НМ, є вдосконалення математичної моделі.
6. Доведена доцільність використання НМ в задачах біометричної ідентифікації суб’єктів.

Після дослідження показників КС була отримана наступна інформація:

Результуючий рівень помилкової ідентифікації становить 9,51%, що

задовольняє поставленим умовам. Збільшення розмірів вибірки, розширення її репрезентативності, покращення апаратних можливостей та зміна політики часових обмежень на навчання НМ дають можливість перевершити отримані результати.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Терейковський І. Нейронні мережі в засобах захисту комп'ютерної інформації / І. Терейковський. К. : ПоліграфКонсалтинг. 2007. – 209 с.
2. Jeffrey L. Elman Finding Structure in Time // COGNITIVE SCIENCE 14, 179-211 (1990)
3. Cimarusti D., Ives R.B. Development of an automatic identification system of spoken languages: Phase 1. In Proceedings IEEE International conference on acoustic, speech and signal processing, Paris, 1982.
4. Задоров В.Б. Системний аналіз об'єктів і процесів: технологічні основи: Навчальний посібник. – К.: КНУБА - 2003. – 276 с.
5. Вакуленко А. Биометрические методы идентификации личности: обоснованный выбор и внедрение / А. Вакуленко, А. Юхин. – М.: Наука, 2007. – 224 с..
6. Вилков А.С. Информационная безопасность персональных ЭВМ и мониторинг компьютерных сетей / А.С. Вилков. – М. : МИНИТ ФСБ России, 2005. – 210 с.
7. Галушкин А. И. Теория нейронных сетей / А. И. Галушкин. М. : ИПРЖР, 2000. 416 с.
8. Горбань А. Н. Обучение нейронных сетей / А. Н. Горбань. М. : ParaGraph, 1990. 160 с/
9. Айвенс К. Компьютерные сети / Айвенс К. ; пер. с. англ. – СПб. : Питер, 2006. – 304 с.
10. Li K.P., Edwards T.J. Statistical models for automatic language identification. In Proceedings IEEE International conference on Acoustic, Speech and Signal Processing 80, Denver, CO, 1980.

11. Leonard R.G., Doddington G.R. Automatic language identification. Technical report RADC-TR-74-200, Air Force Rome Air Development Center, 1974.