

Вінницький національний технічний університет  
Факультет інформаційних технологій та комп'ютерної інженерії

Програмний засіб для розпізнавання  
пошкоджених друкованих текстів із виконанням  
багаторівневого аналізу документа

Виконала: ст. гр. 1КІ-19м Поліщук К. В.

Керівник: д.т.н., проф. Мартинюк Т. Б.

Актуальність створення засобів  
розпізнавання символів

Актуальність роботи:  
подальше вдосконалення методів  
розпізнавання текстових  
документів, у тому числі  
пошкоджених

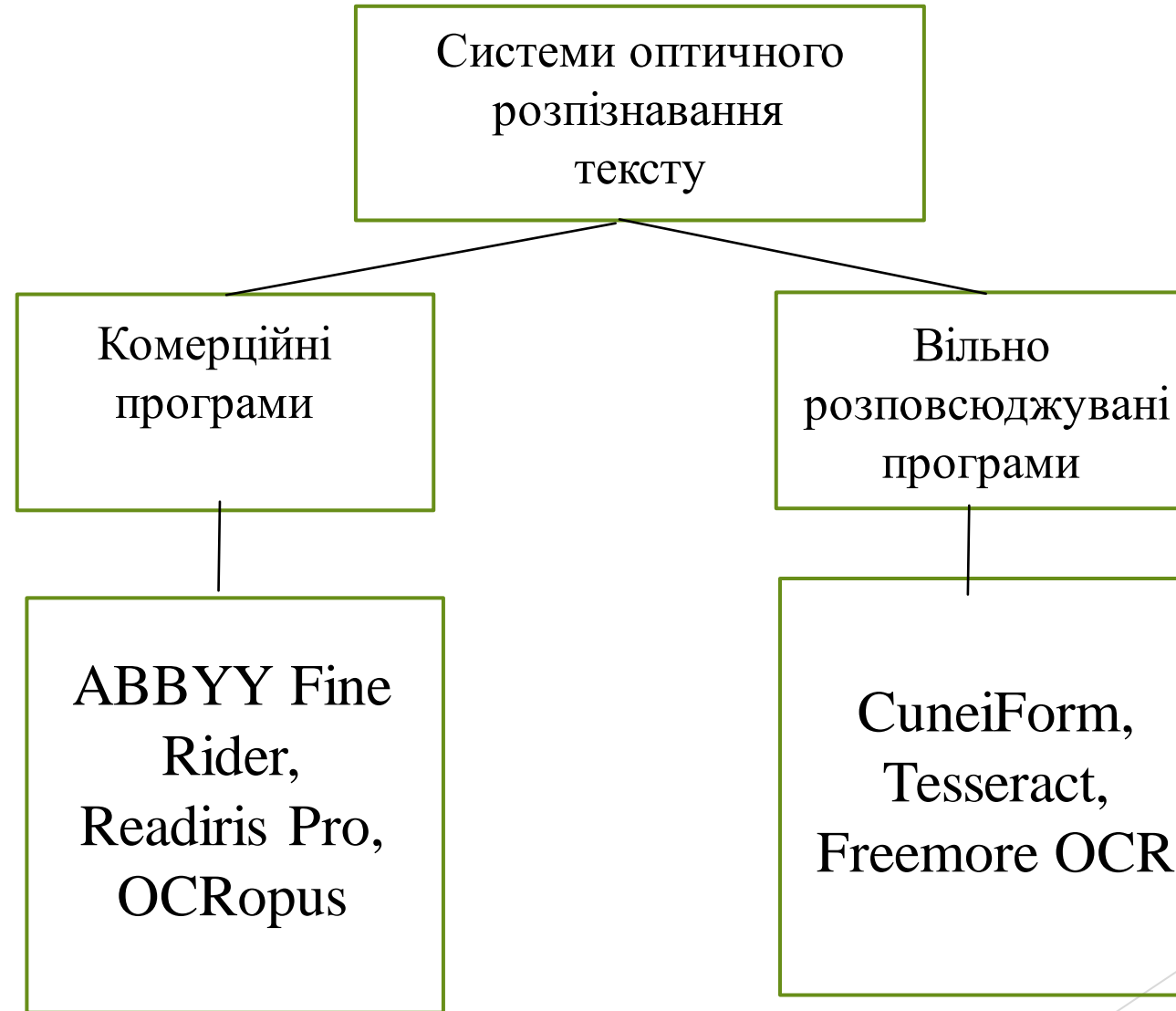
# Завдання роботи по розпізнаванню символів

1. Провести аналіз існуючих методів оброблення цифрового зображення для пошуку та розпізнавання символів текстового документу;
2. Запропонувати поліпшений метод оброблення цифрового зображення текстового документу для виділення та подальшого розпізнавання символів текстового документу;
3. Розробити програму оброблення зображення для розпізнавання символів текстового документу.

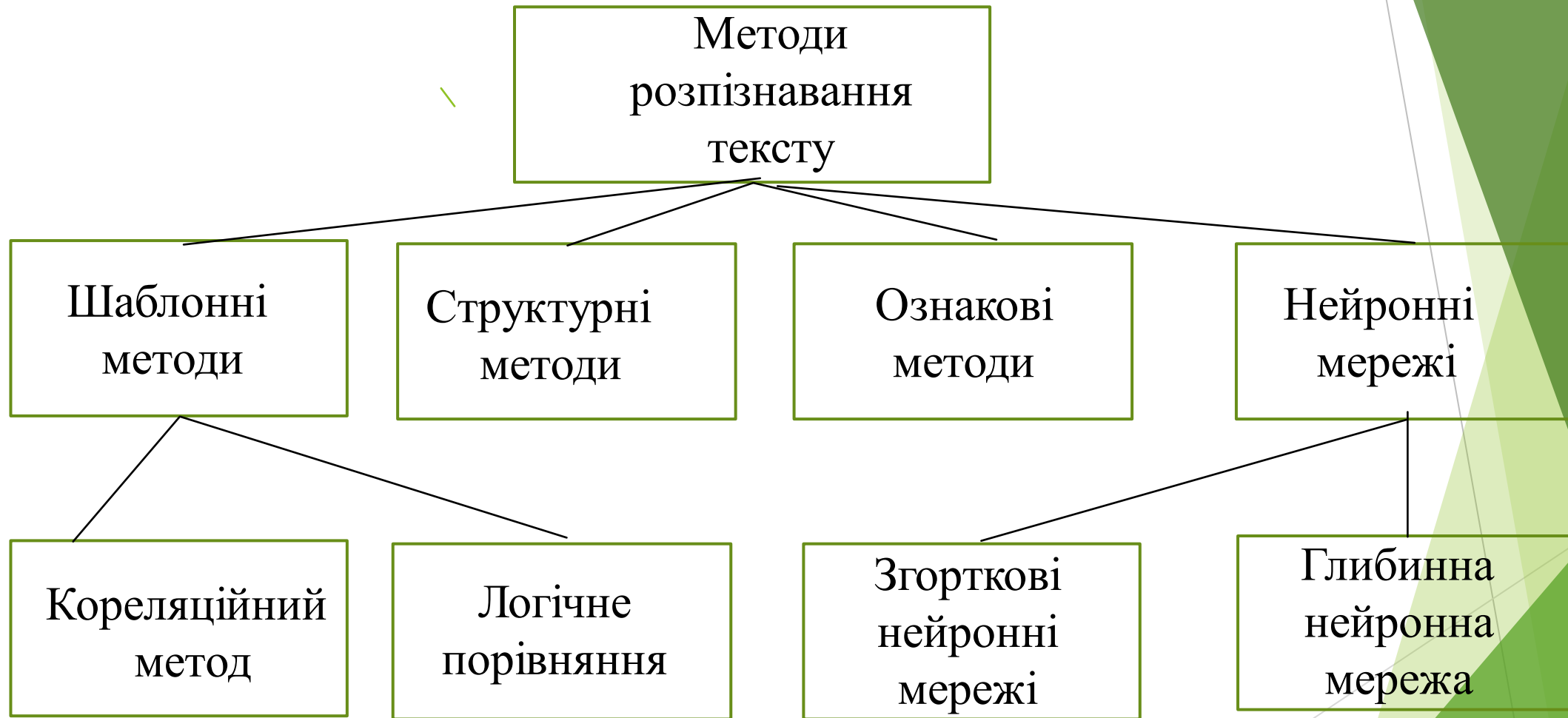
# Дефекти текстових документів



# Системи розпізнавання тексту



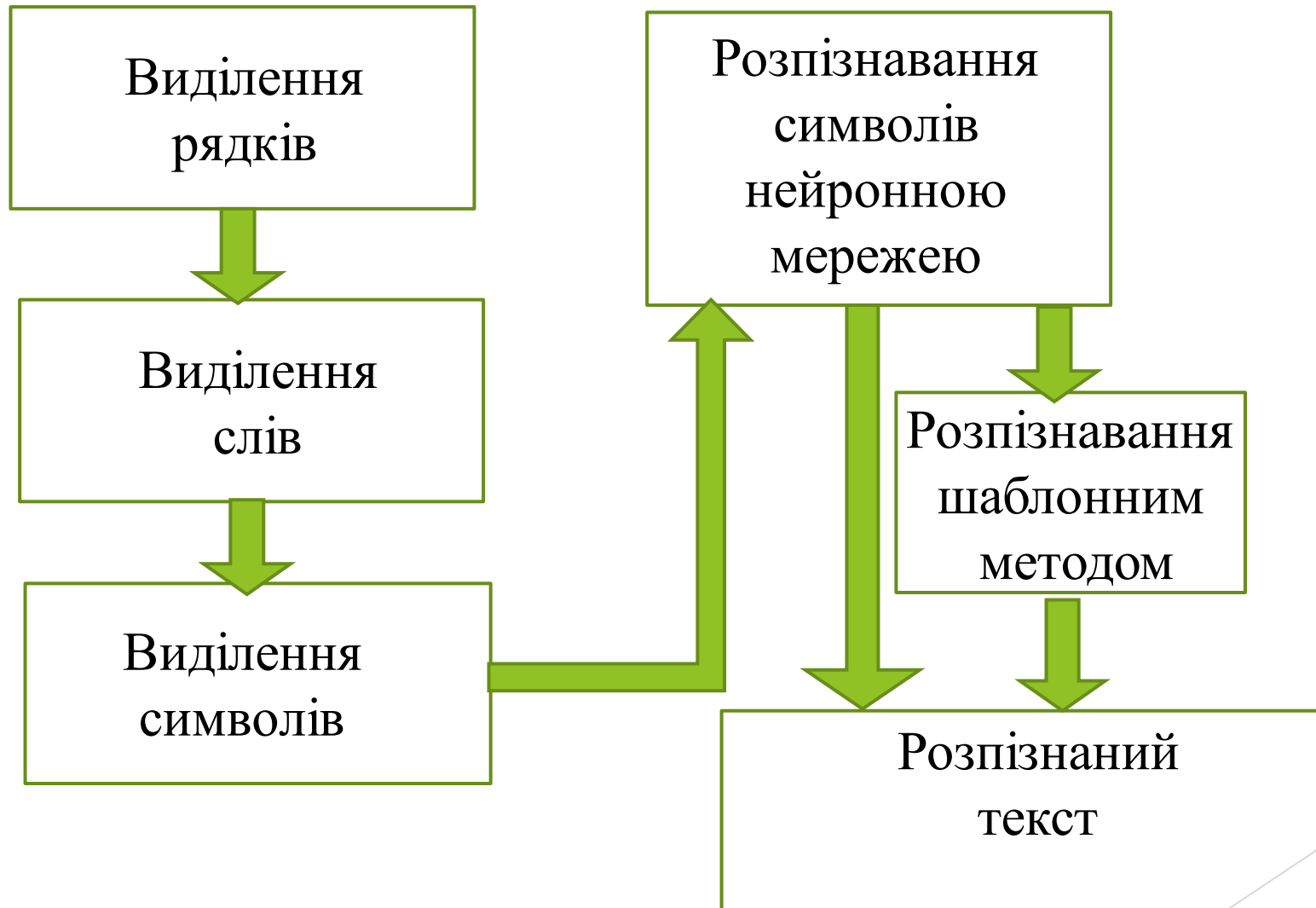
# Методи розпізнавання символів тексту



# Загальна послідовність розпізнавання тексту

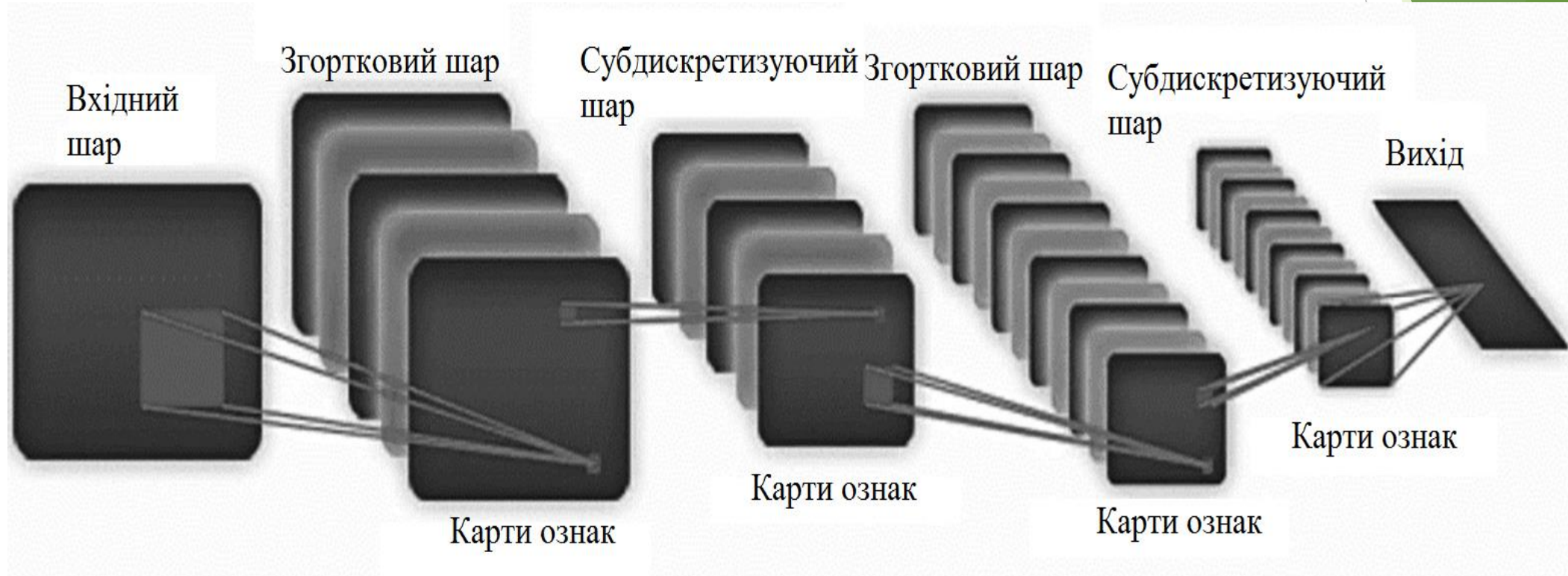


# Послідовність розпізнавання символів тексту





# Нейронна мережа



# Знаходження вірного символу

$S[i,j]$  - символ, що перевіряється

$E[i,j]$  - еталонний символ

$MF[i,j] = S[i,j] \vee E[i,j]$  - матриця фону

$MZ[i,j] = S[i,j] \wedge E[i,j]$  – матриця зображення символу

$K_S = \Sigma (MZ[i,j], \text{якщо } x_{ij} = 1)$  - коефіцієнт збігу символу

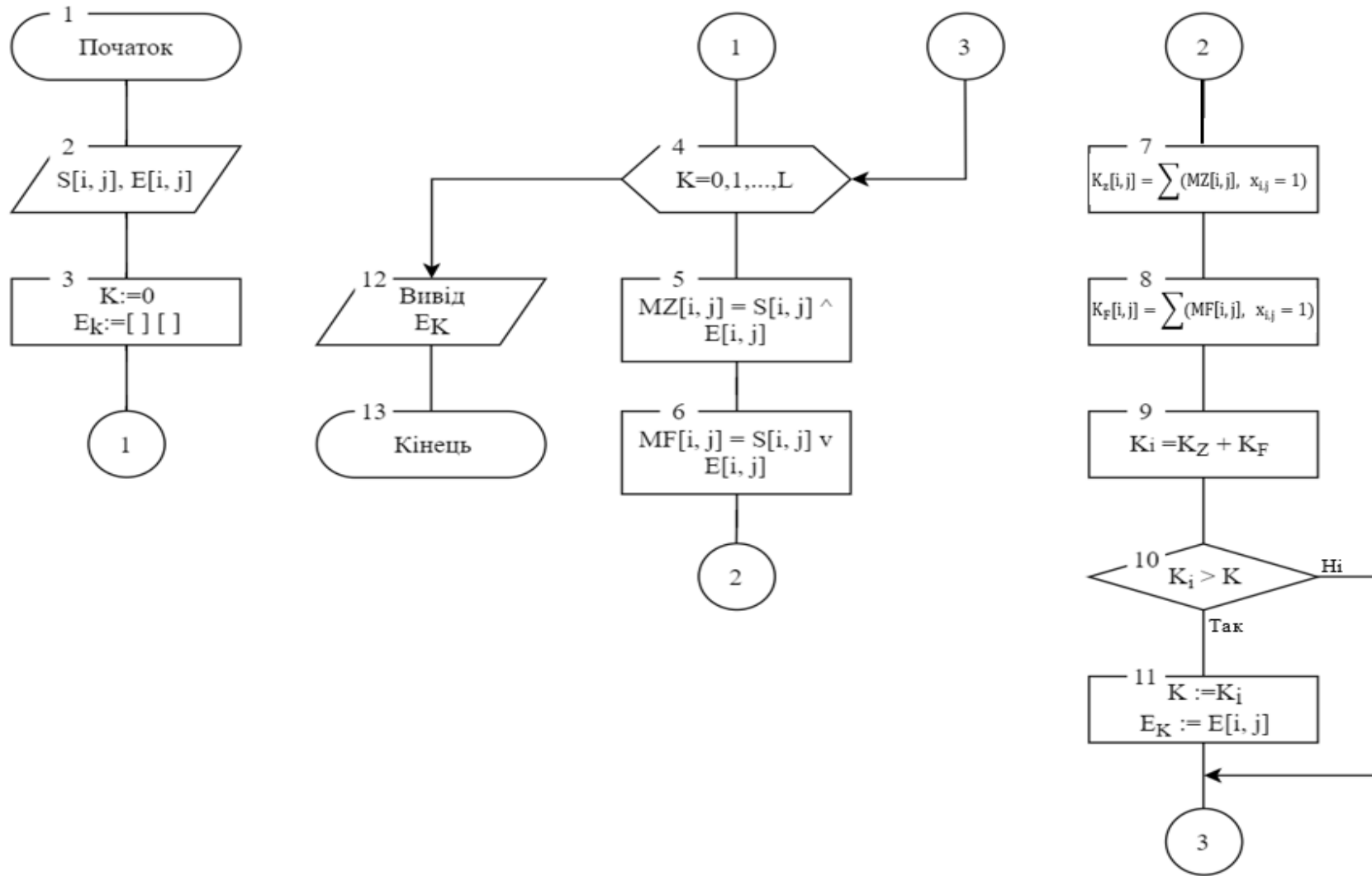
$K_F = \Sigma (MF[i,j], \text{якщо } x_{ij} = 0)$  - коефіцієнт відхилення фону

$$K_i = K_F + K_S$$

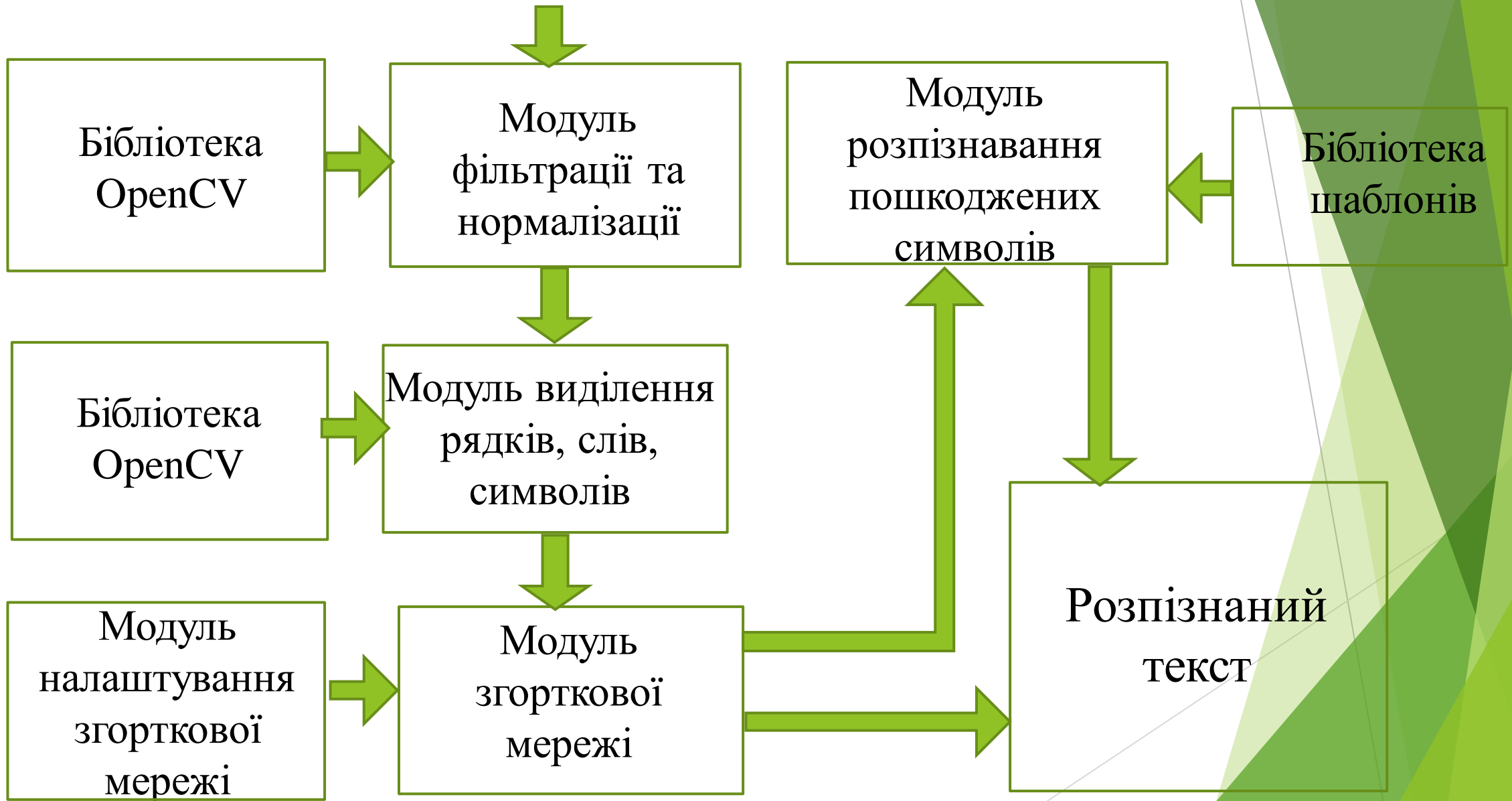
$$K_{\max} = \{K_1, K_2, \dots, K_L\}$$

$$S[i,j] = E[i,j]_{\max}$$

# Алгоритм розпізнавання символів



# Структура програми



# Варіанти пошкоджених символів

СИМВОЛ

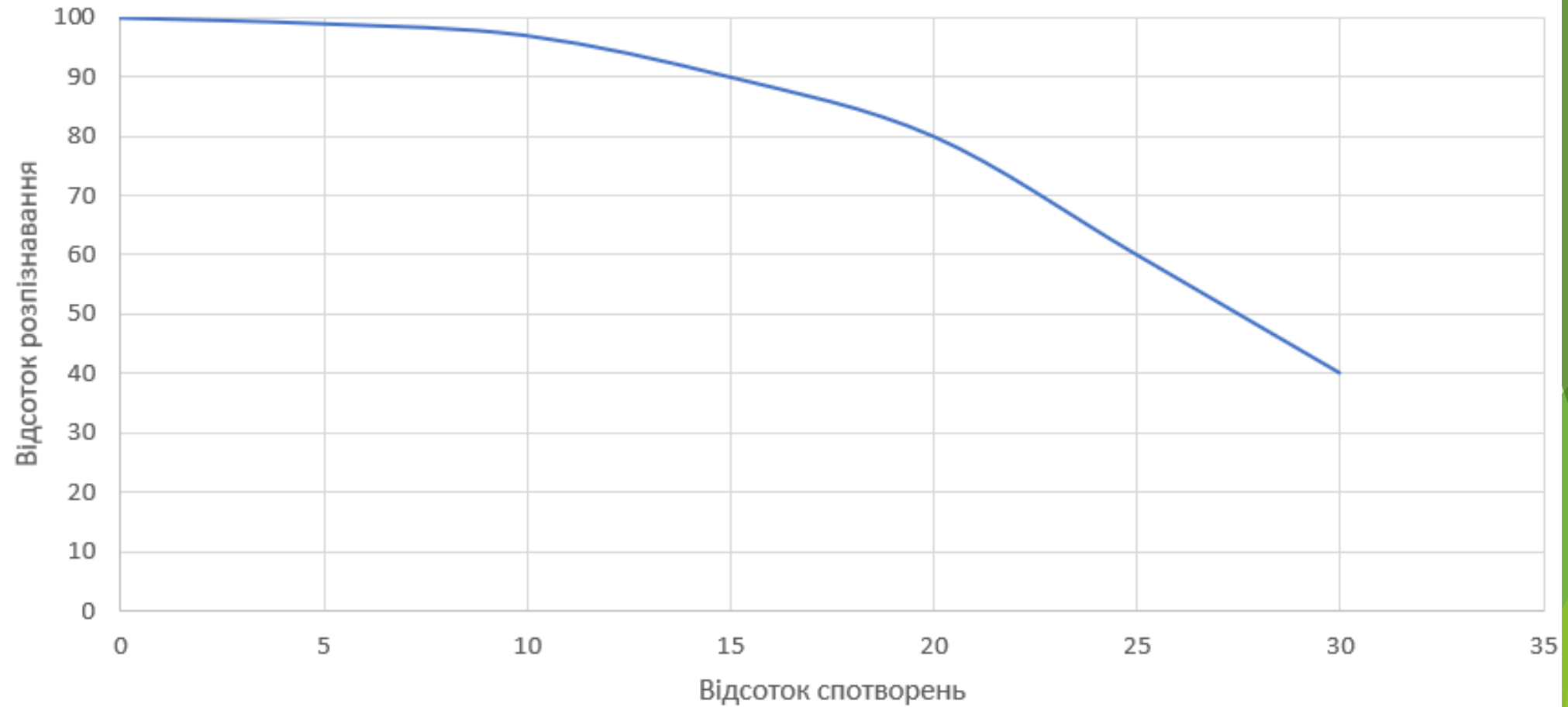
СИМЕСЛ

СИМВОСЛ

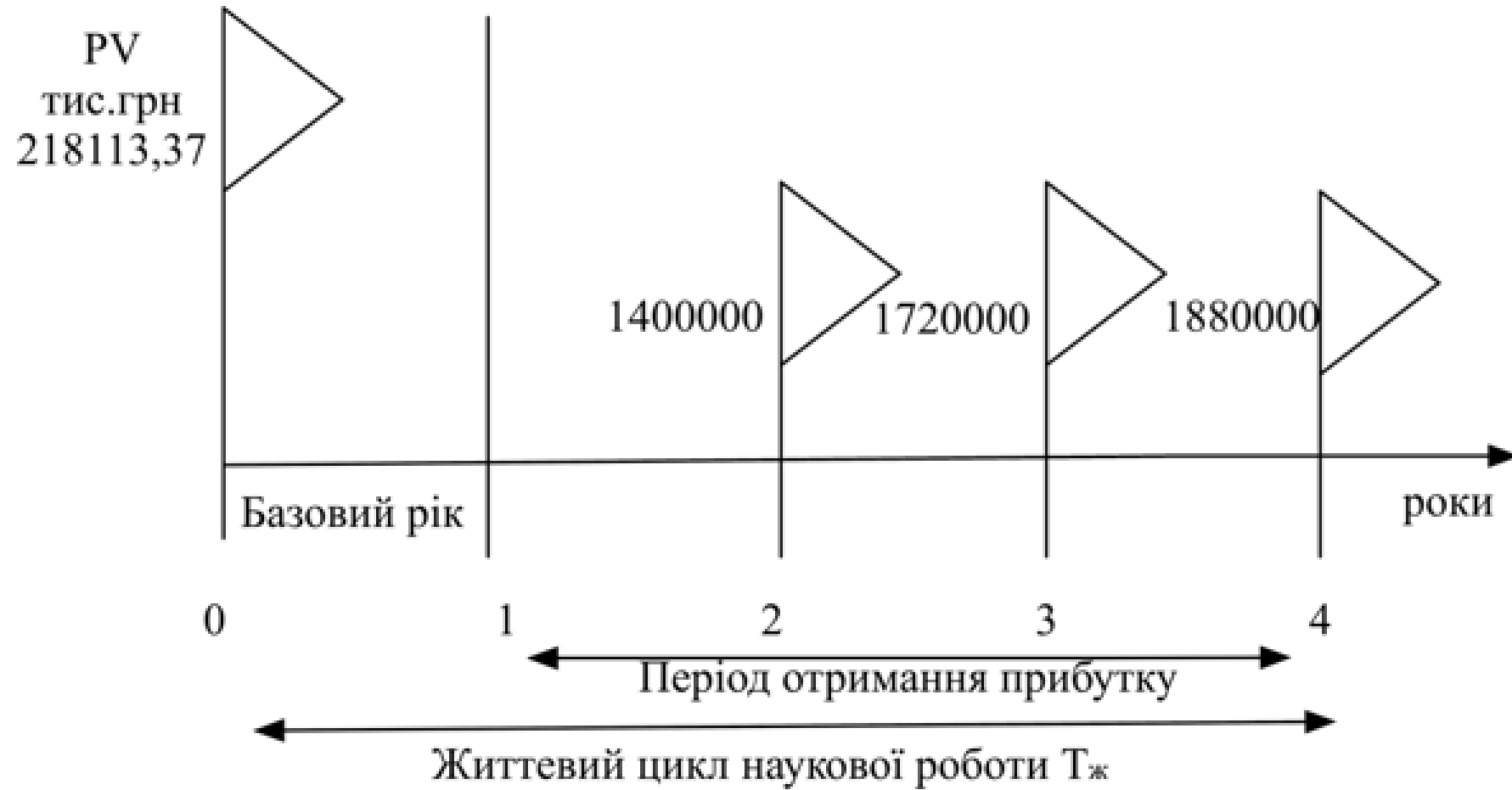
СИМВОЛ

# Результати перевірки ефективності роботи

Графік залежності відсотку розпізнавання символу від його спотворення



# Термін окупності програми



# Новизна роботи

- ▶ удосконалено метод оброблення зображення текстового документу для виділення та розпізнавання символів тексту, який відрізняється від відомих підходів виконанням обчислення коефіцієнтів збігів символу зображення і його фону та їх порівняння із еталонними зображеннями, що дозволяє більш ефективно розпізнавати символи у отриманому пошкодженому друкованому тексті.



# Практичні результати роботи

- ▶ створено алгоритм оброблення цифрових зображень для виділення та розпізнавання пошкоджених символів друкованих текстів;
- ▶ розроблено програму оброблення зображення для виділення символів та їх розпізнавання у пошкоджених друкованих текстах.

# Публікації

- ▶ Поліщук К. В. Розпізнавання деформованих символів текстових документів // Електронна наукова конференція ВНТУ «Молодь в науці: дослідження, проблеми, перспективи (МН-2019)». Режим доступу: <https://conferences.vntu.edu.ua/index.php/mn/mn2019/paper/viewFile/8113/6786>
- ▶ Поліщук К. В. Розпізнавання пошкоджених друкованих текстів із використанням багаторівневого аналізу документа / К. В. Поліщук, М.А.Очкуров // Електронна наукова конференція ВНТУ «Молодь в науці: дослідження, проблеми, перспективи (МН-2021)». Режим доступу: <https://conferences.vntu.edu.ua/index.php/mn/mn2021/paper/viewFile/11034/9190>

Дякую за увагу!

Поліщук К. В.