

InkScan - Aplicație dedicată recunoașterii scrisului de mână

Autor: Dudună Daria-Nicole, 332AB

An: 2025

1. INTRODUCERE

1.1. Prezentarea temei alese

Pentru realizarea proiectului, mi-am propus implementarea unei aplicații dedicate recunoașterii scrisului de mână, bazată pe procesarea imaginilor, ținând cont că, în ziua de astăzi, digitalizarea documentelor reprezintă o necesitate în viața tuturor, dar totodată și un confort.

Scrisul de mână a reprezentat, pe parcursul ultimilor mii de ani, o modalitate de exprimare pentru oameni, principalul mod de a stoca date până la apariția tehnologiei și un mod de a comunica între ei. InkScan dorește adaptarea acestei metode la vremurile de astăzi, combinând lumea digitală cu scrisul de mână autentic și specific omenirii, pentru a ușura activitatea utilizatorilor din diferite domenii.

1.2. Obiective

În contextul actual, stocarea eficientă și accesul rapid al datelor joacă un rol important în numeroase domenii precum educația, cercetarea sau administrarea. Cel puțin în rândul studenților, scrisul de mână reprezintă o modalitate de a procesa mai ușor informația, dar și o modalitate de stocare mai greu accesibilă. Astfel, proiectul InkScan își propune realizarea unei interfețe intuitive, însă totodată complexe, încercând să-i ofere utilizatorului o experiență plăcută în digitalizarea documentelor scrise.

O prioritate o reprezintă accesibilitatea și corectitudinea informației furnizate în urma procesării fișierelor date de utilizatori, așadar, în implementarea aplicației am folosit modele de învățare automată, modele de corectare gramaticală, instrumente de redare audio și tehnologii de recunoaștere optică a caracterelor.

Câteva obiective generale care au stat la baza implementării proiectului:

- Dezvoltarea unei aplicații web care e ușor de folosit
- Implementarea unui sistem robust de detecție și corecție a scrisului de mână
- Integrarea unui mecanism de detecție a orientării scrisului
- Oferirea unor opțiuni de export în funcție de preferința utilizatorului
- Prezentarea unei metode de stocare a informațiilor prin salvarea activității utilizatorilor
- Implementarea unor funcții de redare audio și a unei interfețe dedicată scrisului de mână online

În comparație cu alte aplicații existente, care oferă OCR parțial sau doar pe text tipărit, acest proiect se remarcă prin adaptabilitatea sa la scrisul cursiv (sau de mână), prin procesarea avansată a imaginilor și prin funcții de recunoaștere, corectare, salvare în cadrul aceleiași platforme. De asemenea, InkScan cunoaște și promovează beneficiile scrisului de mână, oferindu-le utilizatorilor posibilitatea să facă chiar acest lucru.

2. SUPORT TEHNIC ȘI REFERINȚE

Pentru a înțelege contextul actual al aplicațiilor care oferă conversia scrisului de mână în text digital, vom analiza soluții reprezentative ale pieței curente, cu scopul de a evidenția modul în care funcționează implementările existente, tehnologiile utilizate, precum și avantajele și limitările acestora.

În continuare, voi prezenta trei aplicații relevante – Transkribus, Google Keep și Microsoft OneNote – care oferă funcționalități similare cu aplicația dezvoltată, urmate de o evaluare obiectivă a fiecăreia. Analiza aceasta va contribui la evidențierea poziționării proiectului actual în raport cu soluțiile deja existente.

2.1. Transkribus – Platformă de recunoaștere a scrisului de mână

Transkribus este una dintre cele mai avansate platforme destinate extragerii textului scris de mână, utilizată în special în domeniul arhivisticii și istoriei documentelor. Aceasta combină tehnologia Handwritten Text Recognition (HTR) cu inteligență artificială, fiind capabilă să recunoască atât texte istorice, cât și moderne scrise de mână.

Conform studiului [1] Transkribus se bazează pe rețele neuronale recurente (RNN) și long short-term memory (LSTM) pentru a învăța modele de scris, permițând o acuratețe crescută în transcrierea documentelor, chiar și în condiții de degradare fizică a acestora, de aceea fiind potrivite pentru texte istorice. Pentru antrenarea modelului au fost folosite pagini transcrise manual, modelul Universal Giant fiind antrenat cu peste 15 milioane de cuvinte, în diferite tipuri și stiluri de scris de mână. [2]

Puncte tari:

- Acuratețe ridicată în recunoașterea scrisului de mână
- Acces la modele pre-antrenate
- Interfață grafică intuitivă

- Opțiune de a personaliza un model AI pe scrisul utilizatorului.

Puncte slabe:

- Necesită mult efort inițial pentru transcrierea și corectarea datelor de antrenament, ceea ce poate încetini procesul de automatizare. Conform
- Performanță scăzută în cazul scrisului de mână foarte dezorganizat sau cu multe corecturi.
- Accesibilitate limitată în planul gratuit.

2.2. Google Keep – Recunoașterea scrisului de mână în aplicații mobile

Google Keep este o aplicație dedicată notițelor dezvoltată de Google, care oferă funcționalități de recunoaștere optică a caracterelor pentru extragerea textului din imagini, inclusiv cel scris de mână. Această funcționalitate permite utilizatorilor să fotografieze notițele scrise manual și să le convertească în text editabil direct în aplicație. [3]

Google Keep a introdus și un mod dedicat pentru notițe scrise de mână, separat de instrumentele de desen existente. Acest mod permite utilizatorilor să interacționeze cu segmente de text, oferind opțiuni precum "copy as text" pentru a extrage doar porțiunile dorite dintr-un fragment scris manual.

De asemenea, Google Keep utilizează API-ul Google Cloud Vision pentru detectarea și extragerea textului scris de mână din imagini, care oferă tehnici avansate de recunoaștere a textului, inclusiv pentru documente dense permițând extragerea structurată a informațiilor din imagini. Google Keep poate detecta până la 97 de limbi. [4]

Puncte tari:

- Integrare în cadrul ecosistemului Google: Permite salvarea și partajarea ușoară a notițelor în Google Docs de exemplu.
- Accesibilitate: Disponibilitate largă, facilitând accesul la notițe de pe diverse dispozitive.
- OCR eficient: Capabilă să extragă text din imagini cu o acuratețe ridicată.

Puncte slabe:

- Limitări: Pentru scrisul de mână dezordonat sau cu stiluri variate, acuratețea OCR poate scădea semnificativ.
- Nu permite utilizatorilor să antreneze modele OCR adaptate la propriul stil de scris.

- Nu oferă sugestii de corectare a textului transcris

2.3. Microsoft OneNote – Funcționalitatea „Ink to Text”

Microsoft OneNote oferă o funcționalitate pentru conversia scrisului de mână în text digital, denumită „Ink to Text”. Această caracteristică permite utilizatorilor să scrie manual folosind un „stylus” sau degetul și să convertească ulterior notițele în text editabil, facilitând astfel organizarea și partajarea informațiilor.

Funcționalitatea „Ink to Text” se bazează pe recunoașterea optică a caracterelor (OCR) integrată în OneNote. Utilizatorii pot selecta textul scris de mână folosind instrumentul „Lasso Select” și apoi aplica opțiunea „Ink to Text” . [5]

Recent, Microsoft a introdus și instrumentul „Ink to Text Pen”, care permite conversia în timp real a scrisului de mână în text digital pe măsură ce utilizatorul scrie. Această funcționalitate este disponibilă în aplicații precum Word, PowerPoint și OneNote pentru Windows și oferă posibilitatea de editare a textului prin gesturi - cum ar fi ștergerea sau inserarea de linii noi.

Puncte tari

- Conversie precisă: OneNote oferă o conversie precisă a scrisului de mână în text, facilitând citirea și editarea notițelor.
- Integrare cu alte aplicații Microsoft: Notițele convertite pot fi ușor integrate în alte aplicații din suita Microsoft 365

Puncte slabe

- Disponibilitate limitată: Funcționalitatea „Ink to Text” nu este disponibilă în toate versiunile , ceea ce limitează accesibilitatea pentru unii utilizatori.
- Necesitatea unui dispozitiv compatibil: Pentru a utiliza funcționalitatea în timp real, este necesar un dispozitiv compatibil cu inputul tactil sau un „stylus”, ceea ce poate reprezenta o barieră pentru unii utilizatori.
- Lipsa personalizării modelelor de recunoaștere: OneNote nu permite antrenarea de modele personalizate pentru recunoașterea scrisului de mână, ceea ce poate afecta acuratețea pentru stiluri de scris ieșite din comun.

- Limitări în recunoașterea limbilor: Deși suportă mai multe limbi, funcționalitatea poate avea limitări în recunoașterea precisă a scrisului de mână în limbi cu caractere speciale sau diacritice.

Aplicațiile analizate mai sus oferă funcționalități relevante pentru recunoașterea scrisului de mână, fiecare cu propriile puncte forte și limite. Transkribus se remarcă prin acuratețea în contexte istorice, dar este complex și greu de personalizat; Google Keep oferă o experiență simplificată și accesibilă, dar cu limitări în precizia recunoașterii; OneNote se distinge prin integrare în ecosistemul Microsoft și suport extins pentru stilouri digitale, însă este dependent de infrastructura comercială. Aceste observații confirmă direcția și relevanța obiectivelor stabilite în proiectul nostru, subliniind nevoia unei aplicații personalizabile, cu procesare locală, care să combine OCR avansat, preprocesare dinamică a imaginilor și interfață prietenoasă. În continuare, vor fi detaliate fundamentele teoretice ce susțin această abordare complexă, de la procesarea imaginilor, la tehnici de segmentare, aliniere și sinteză vocală.

Context teoretic

Formatul JPG (sau JPEG) este cel mai folosit atunci când vine vorba de imagini. E bun pentru fotografii, pentru că reduce dimensiunea fișierului prin comprimare cu pierderi, adică „taie” din detalii pentru a ocupa mai puțin spațiu. Calitatea poate scădea, dar pentru uz general e suficient de bun.

Formatul PNG, în schimb, păstrează toate detaliile originale pentru că folosește o comprimare fără pierderi. Reprezintă alegerea potrivită când e nevoie de o anume claritate, cum ar fi la logo-uri, grafice sau imagini cu fundal transparent.

Flask este un framework Python pentru aplicații web. Permite conectarea frontendului cu backendul, putând implementa diverse funcționalități precum autentificarea sau conectarea la o baza de date.

3.IMPLEMENTAREA PROIECTULUI

3.1. Arhitectura generală a aplicației

Aplicația InkScan este structurată sub forma unei aplicații web dezvoltate cu ajutorul framework-ului Flask, în Python. Aceasta este organizată în 6 fișiere html + javascript, 5 fișiere css și un fișier python. Pentru stocarea datelor utilizatorilor am folosit un fișier JSON.

Aplicația are la bază numeroase aspecte teoretice legate de multimedia și modele de învățare automată, în implementarea sa fiind incorporate o multitudine de biblioteci și metode, care vor fi detaliate în continuare.

3.2. Biblioteci, API-uri și framework-uri folosite

- Flask: framework-ul principal pentru crearea serverului web și pentru maparea paginilor.
- TrOCR (Transformers) - model de tip encoder-decoder bazat pe Vision Transformer pentru recunoașterea scrisului de mână.
- EasyOCR – model de învățare automată utilizat pentru obținerea coordonatelor exacte ale cuvintelor din imagine și a scorurilor de încredere.
- Pytesseract - utilizat pentru estimarea unghiului de rotație al textului (OSD).
- OpenCV (cv2) și imutils - pentru procesarea imaginilor: conversii, rotații, aliniere, diverse efecte..
- PIL (Pillow) - pentru manipularea imaginilor (watermark).
- TextBlob, WordNinja, LanguageTool - pentru procesarea și corectarea automată a textului.
- ReportLab, python-docx - pentru generarea fișierelor PDF și DOCX din textul recunoscut.

3.3. Algoritmi și metodologie. Etape

- **Autentificare(Login):** Utilizatorul introduce datele contului său care sunt validate sau nu prin verificarea acestora în fișierul „users.json”. În cazul în care acesta nu are un cont existent, își poate crea unul prin metoda de Înregistrare(Register), datele introduse fiind scrise apoi în fișierul json.
- **Preprocesare imagine:** Utilizatorul încarcă în aplicație o imagine în format jpg, png sau jpeg, iar această imagine este apoi transmisă către backend. Apoi aceasta este preprocesată, aplicând un filtru alb-negru folosind metoda `cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)` care returnează imaginea în grayscale (adică având un singur canal). Pentru a trece de la 3 canale la unul singur, se atribuie o pondere fiecărui canal, valorile pixelului se înmulțesc fiecare cu ponderea asociată și se adună rezultând valoarea pixelului în spațiul de culoare gray. Totuși metodele OCR trebuie să primească o imagine în format RGB, așadar vom face conversia înapoi în acest spațiu de culoare, folosind aceeași funcție din biblioteca cv2, dar cu argumentul `cv2.COLOR_GRAY2RGB`. Preprocesarea aceasta are rolul de a uniformiza imaginea pentru o detecție mai ușoară a conținutului și pentru a elimina zgomotul.
- **Corectarea orientării imaginii:** La acest pas, aplicația se asigură că imaginea este orientată corespunzător, adică că textul din imagine este aliniat atât pe verticală, cât și pe orizontală, pentru

a asigura funcționarea corectă a sistemului de recunoaștere. Imaginea este citită folosind metoda `cv2.imread(path)` în format BGR, se face apoi conversia în spațiul de culoare RGB cu ajutorul metodei menționate la pasul anterior și cu argumentul `cv2.COLOR_BGR2RGB` unde doar se schimbă ordinea canalelor pentru fiecare pixel. Apoi se încearcă determinarea unghiului cu care trebuie rotită imaginea în funcție de tendința de orientare a textului cu ajutorul metodei `pytesseract.image_to_osd`. Tesseract identifică zonele cu text și folosește analiza statistică asupra direcției liniilor drepte pe care le interpretează că ar forma cuvinte. Dacă cuvintele par scrise vertical, el va detecta că textul este rotit cu 90 de grade. Componenta OSD (Orientation and Script Detection) returnează acest unghi. Totuși, aceasta metodă nu asigură o rotație perfectă, având unele limitări – adică metoda nu poate detecta dacă poza este cu capul în jos sau nu, ci doar dacă orientarea blocurilor detectate (cuvintelor) este bună. Așadar, pentru a face distincția dintre cele 2 cazuri, vom analiza mai departe atât imaginea corectată, cât și varianta sa rotită cu 180 de grade. Aceste rotații sunt făcute cu ajutorul metodei `rotate_bound` din biblioteca `imutils`. Apoi vom încerca detecția cuvintelor folosind EasyOCR pentru ambele imagini și se compară scorurile confidence returnate. În funcție de scorul mai mare, vom ști care imagine este cea orientată corespunzător. EasyOCR este un model de învățare automată bazat pe rețele neuronale. Acesta detectează cadranele în care sunt cuprinse cuvintele, fiecare fiind decupat și trimis către modelul de recunoaștere secvențială: După prelucrare, modelul întoarce textul prezis caracter cu caracter, fiecare rezultat venind cu un scor de încredere între 0 și 1. Pentru comparație, vom face pentru fiecare caz o medie a acestor scoruri.

Pentru rigurozitate, mai departe am implementat o funcție de detecție a înclinării care, spre deosebire de cea precedentă, este destinată unghiurilor mici. Convertim imaginea în tonuri de gri, o netezim cu funcția `GaussianBlur` din `cv2` pentru a reduce zgomotul și a ajuta la o detecție mai clară a conturilor (fiecare pixel din imagine este înlocuit cu o medie ponderată a vecinilor săi), detectăm muchiile (conturile menționate anterior) cu funcția `Canny` din aceeași bibliotecă (care observă schimbarea în intensitatea pixelilor după aplicarea zgomotului gaussian și păstrează cele mai puternice margini), transformăm imaginea în spațiu Hough (spațiu polar) cu ajutorul metodei `HoughLines` și detectăm linii drepte pe baza unghiurilor și a distanțelor față de origine – locul unde se acumulează cele mai multe linii. Construim o matrice de rotație și aplicăm rotația în jurul centrului sau, pentru a alinia textul perfect orizontal.

- **Decuparea și scalarea fiecărui cuvânt și recunoașterea acestuia cu TrOCR:** Cuvintele segmentate de EasyOCR sunt decupate și redimensionate, pentru a le transmite către TrOCR, care în urma analizei are performanțe mai ridicate la partea de recunoaștere a textului scris de mână. Când aceste valori sunt returnate, se reconstruiesc liniile – în funcție de pozițiile detectate de EasyOCR și apoi se reconstruiește textul. Totuși, este de menționat că acest model este sensibil la zgomot și poate detecta caractere false. Astfel, aplicăm mai departe o funcție de curățare a textului, eliminând punctele care sunt plasate greșit, spațiile duble. Apoi, verificăm dacă cuvintele au fost interpretate greșit încercând să despărțim cuvintele mari și să verificăm dacă acestea există în

dictionarul englez. După acest pas, folosim un model hibrid de corectare automata a textului din punct de vedere gramatical si semantic.

- **Generarea documentelor ce pot fi exportate:** textul extras la pasul anterior este acum scris într-un fișier la alegere(docx sau pdf) folosind bibliotecile ReportLab si python-docx, după ce a fost formatat.

- **Funcționalitate audio:** se poate reda textul recunoscut folosind SpeechSynthesis, o interfață JavaScript din Web Speech API, care permite redarea textului recunoscut cu ajutorul motorului de sinteză vocală al browserului.

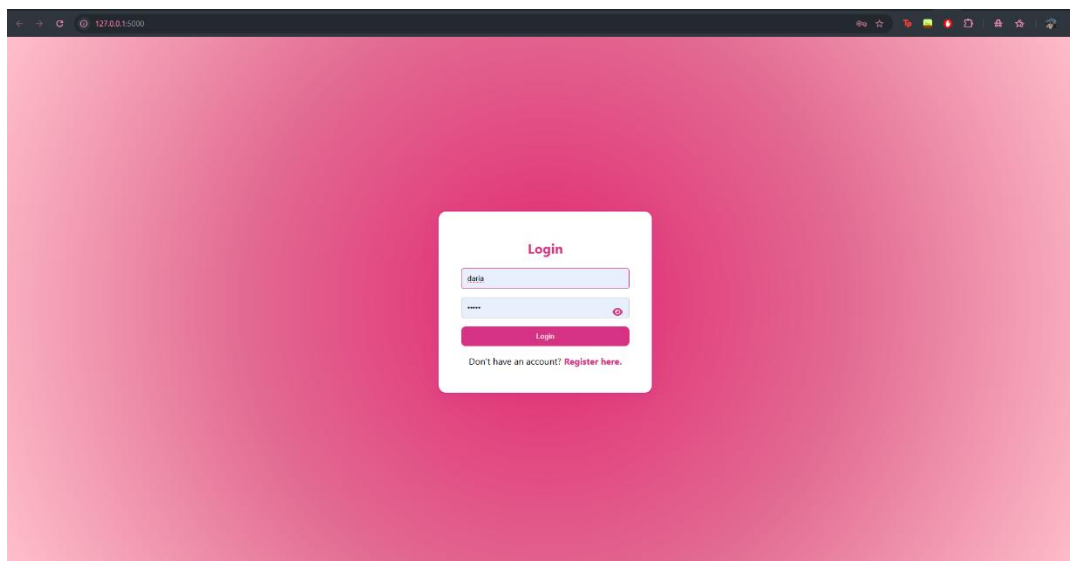
- **Salvarea imaginii desenate cu watermark personalizat si recunoașterea:** Am declarat o tabla in javascript unde utilizatorul poate scrie cu diferite culori. Prin apăsarea butonului de Recognize, tabla la momentul actual este preluată ca text codificat in base64, trimisă către backend in format json. Apoi aceasta este extrasa din string si, cu ajutorul funcției base64.b64decode, este transformata in bytes. Apoi imaginea este transformata in rgb si trimisa către funcția de recunoaștere. In cazul în care se dorește salvarea desenului, se creează un watermark compus din logo-ul aplicației si numele aplicației si cu ajutorul metodelor din biblioteca pillow este poziționat pe imaginea primita cu tabla. Watermark-ul marchează imaginea ca fiind creată sau procesată de InkScan, fiind o modalitatea de a sugera drepturile de autor si de a preveni utilizarea neautorizată sau redistribuirea acesteia fără a se menționa sursa.

- **Vizualizarea si ștergerea istoricului:** Pagina de istoric permite utilizatorului să vizualizeze sau să șteargă anumite părți ale istoricului lor de scanare, metodele fiind bazate pe citirea datelor din fișierul JSON și ștergerea acestora la alegere.

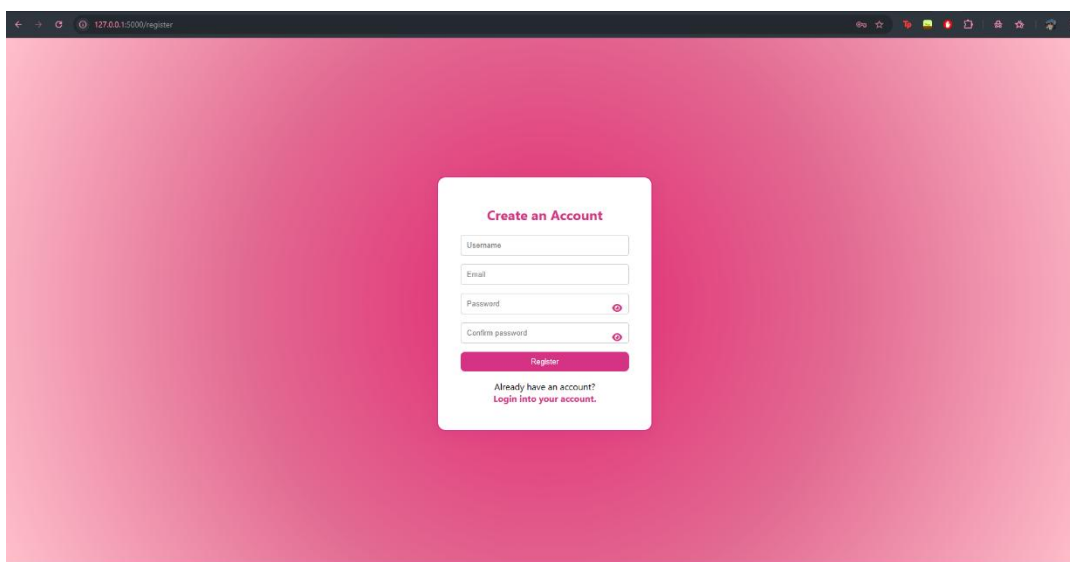
4.MOD DE UTILIZARE

Conform obiectivelor propuse, aplicația implementată este intuitivă și ușor de folosit.

Inițial la pornirea aplicației, este afișată pagina de autentificare(Login). Pentru a utiliza aplicația InkScan, utilizatorul trebuie să își creeze cont, dacă nu are unul existent, folosind pagina Register și introducându-și datele în formular. Prin apăsarea butonului Register, contul este creat și acesta se poate înregistra.



Figură 1 - Pagina de autentificare (Login)



Figură 2 - Pagina de înregistrare (Register)

După validarea contului și înregistrare, utilizatorul este întâmpinat de o pagină intermediară, care conține detalii despre aplicația InkScan și care are un buton „Get started” ce, odată apăsă, redirecționează utilizatorul către pagina principală

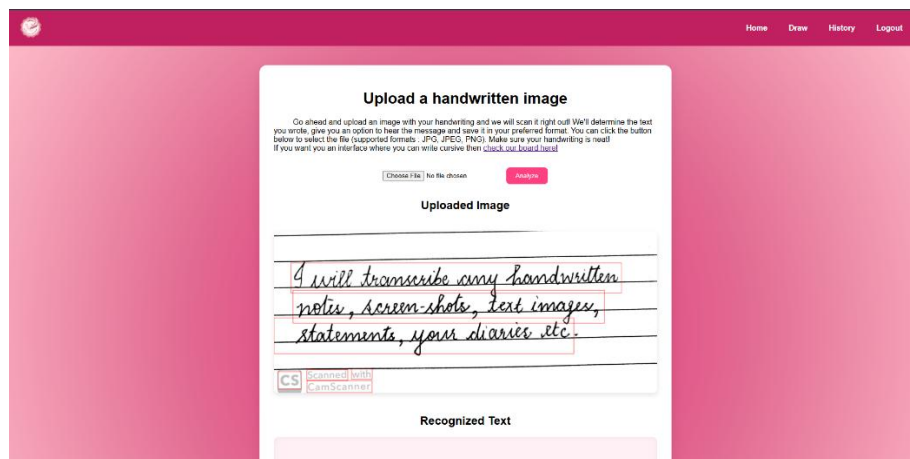
Pagina principală conține un buton pentru încărcarea unui fișier de tip .png, .jpg sau .jpeg. Odată ales acest fișier, butonul Analyze trimite fișierul către backend, iar utilizatorul așteaptă câteva momente procesarea acestuia. Când rezultatul este obținut, pe ecran vor fi afișate imaginea prelucrată – care va avea încadrate cuvintele detectate de aplicație, și textul extras.

Sub textul extras există 2 butoane:

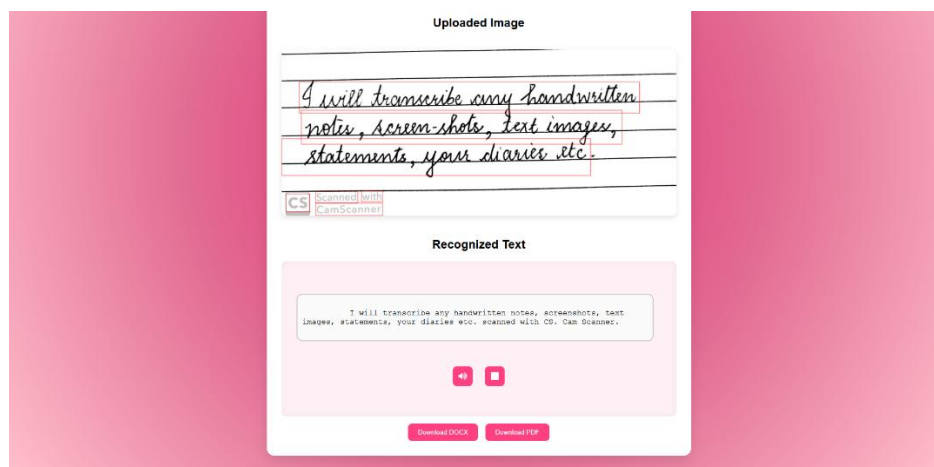
- primul: începe redarea audio a mesajului

- al doilea: oprește redarea audio a mesajului

Sub aceste 2 butoane, utilizatorului îi este prezentată opțiunea de a salva rezultatul scanării sub formă docx sau pdf.



Figură 1 - Pagina principală după atașarea fișierului , se poate observa imaginea prelucrată

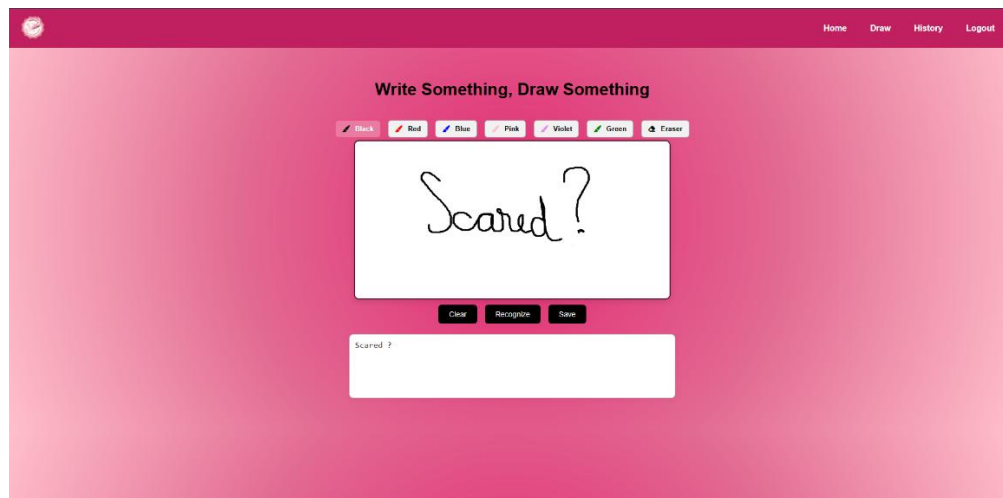


Figură 2 - Continuarea rezultatului din figura 3, ce conține textul extras, butoanele pentru redarea audio și cele pentru salvare

În headerul paginii, observăm bara de navigare, de unde utilizatorul poate merge către o altă pagină sau ieși din cont. Această bară de navigare este prezentă și în celelalte 2 pagini.

Pagina Draw este pagina dedicată interfeței speciale pentru scrisul de mână. Aceasta constituie dintr-o tablă albă, un set de culori si o opțiune pentru ștergere, unde utilizatorul poate scrie cu ajutorul mouse-ului sau creionului, sau poate chiar desena dacă dorește. Culoarea aleasă este sugerată printr-un efect.

După introducerea textului, butonul Recognize poate fi apăsat pentru extragerea mesajului scris. Acest mesaj va apărea în chenarul alb. Butonul Clear golește toată tabla, iar butonul Save duce la salvarea tablei curente ca imagine.



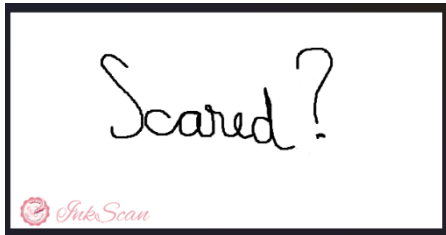
Figură 3 - Pagina cu interfață pentru a scrie

Pagina History este dedicată vizualizării istoricului și păstrării informației într-un loc sigur. Aceasta conține toate fișierele introduse de către utilizator și rezultatele obținute în urma procesării, dar și data scanării. Utilizatorul are opțiunea de a șterge orice fișier folosind butonul Delete din dreptul său, sau de a goli tot istoricul, prin apăsarea butonului Clear History. Ambele operațiuni cer confirmarea utilizatorului printr-o alertă.

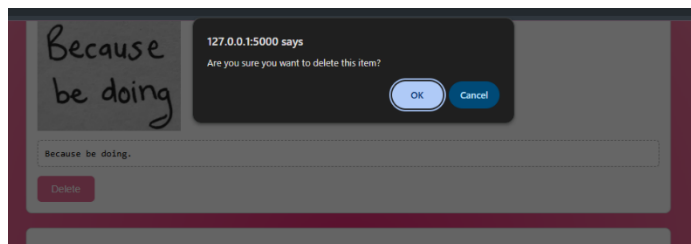


Figură 4 - Pagina History

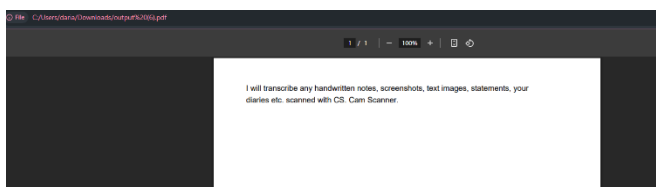
Mai jos voi atașa alte detalii menționate:



Figură 7 - Imaginea salvată cu watermark



Figură 8 - Alertă pentru confirmarea ștergerii unui fișier



Figură 9 - Fișierul salvat pdf

5.CONCLUZII

În dezvoltarea aplicației am încercat mai multe metode pentru a construi o interfață plăcută, pentru a construi un sistem robust care detectează scrisul de mână cu eroare mică.

Aplicația InkScan este o aplicație cu mai multe funcționalități, în principal targetată către studenți, cu scopul de a ușura viața acestora, dar și a oricărui utilizator. Astfel, prin funcția de redare audio a mesajului, studenții pot beneficia de un curs audio, ceea ce le poate spori învățatul. Totodată reprezintă și un loc unde pot păstra aceste fișiere.

Deși, de-a lungul proiectului, acuratețea sistemului a continuat să crească, există niște limitări în ceea ce privește lizibilitatea scrisului pentru a putea fi interpretat, stilurile mai atipice nefiind procesate ca un scris perfect de mână.

Totuși consider că aplicația și-a atins obiectivele propuse, deși există loc pentru îmbunătățire. Față de piața actuală, putem spune ca este o interfață care le cuprinde pe toate – atât de a analiza fișiere text, cât și de a oferi utilizatorului modalitatea a se exprima chiar aici și acum.

6. REFERINȚE

- [1] J. G. P. A. S. e. a. Nockels, "Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research.," Arch Sci 22,, 2022.
- [2] „Transkribus,” [Interactiv]. Available: <https://www.transkribus.org/handwriting-to-text>. [Accesat 6 Mai 2025].
- [3] C. Powell, „Use Google Keep to transcribe your handwritten notes,” 22 August 2017. [Interactiv]. Available: <https://medium.com/@mnmltek/tech-tip-091-use-google-keep-to-transcribe-your-handwritten-notes-fa9e8c18a464>. [Accesat 6 Mai 2025].
- [4] T. D. H. A. R. L. -L. W. a. V. C. D. Keysers, „Multi-Language Online Handwriting Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, nr. 6, pp. 1180-1194, 1 Iunie 2017.
- [5] „Microsoft,” [Interactiv]. Available: <https://support.microsoft.com/en-us/office/take-handwritten-notes-in-onenote-0ec88c54-05f3-4cac-b452-9ee62cebbd4c>. [Accesat 6 Mai 2025].