

Documentatie

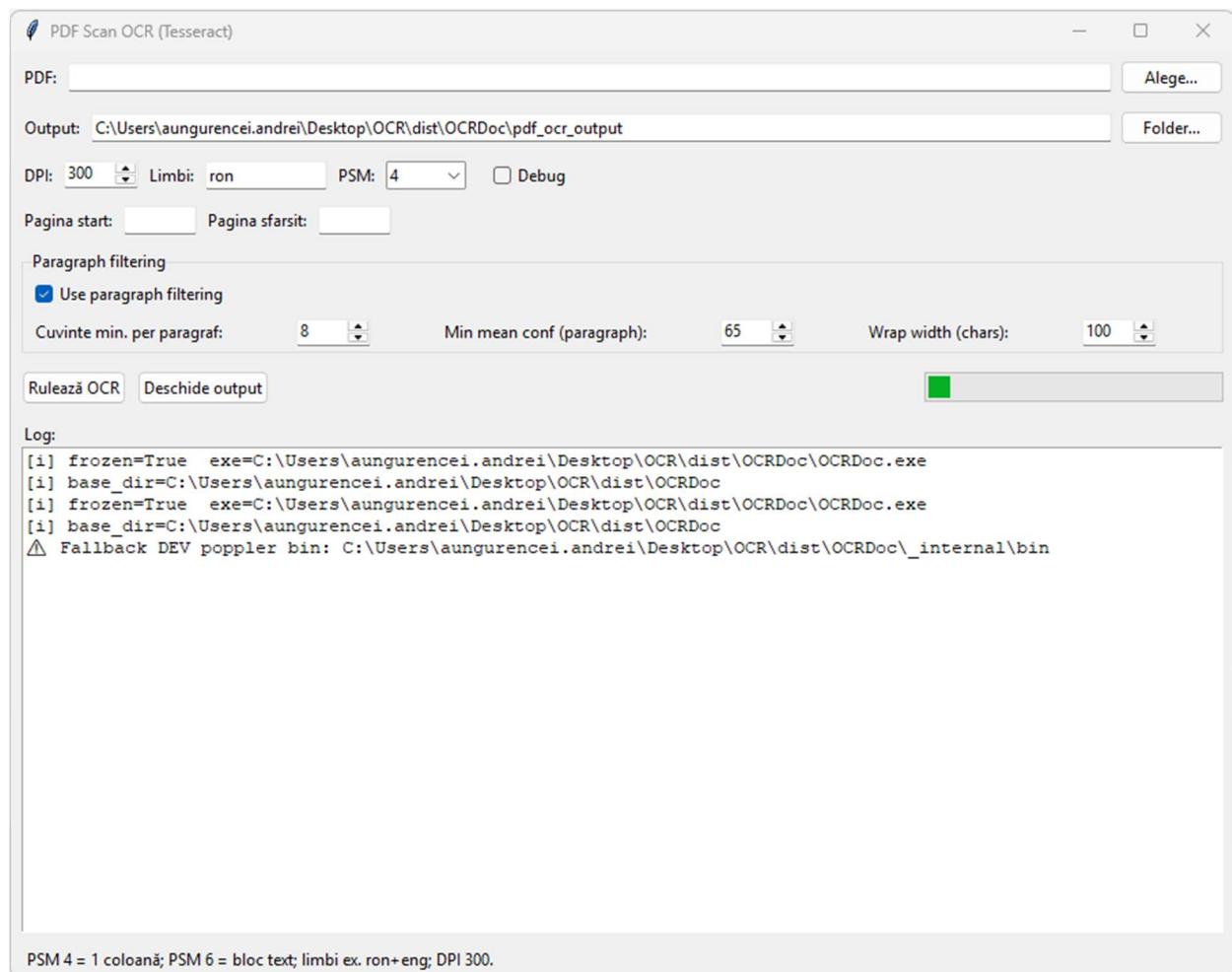
Aplicatie OCR

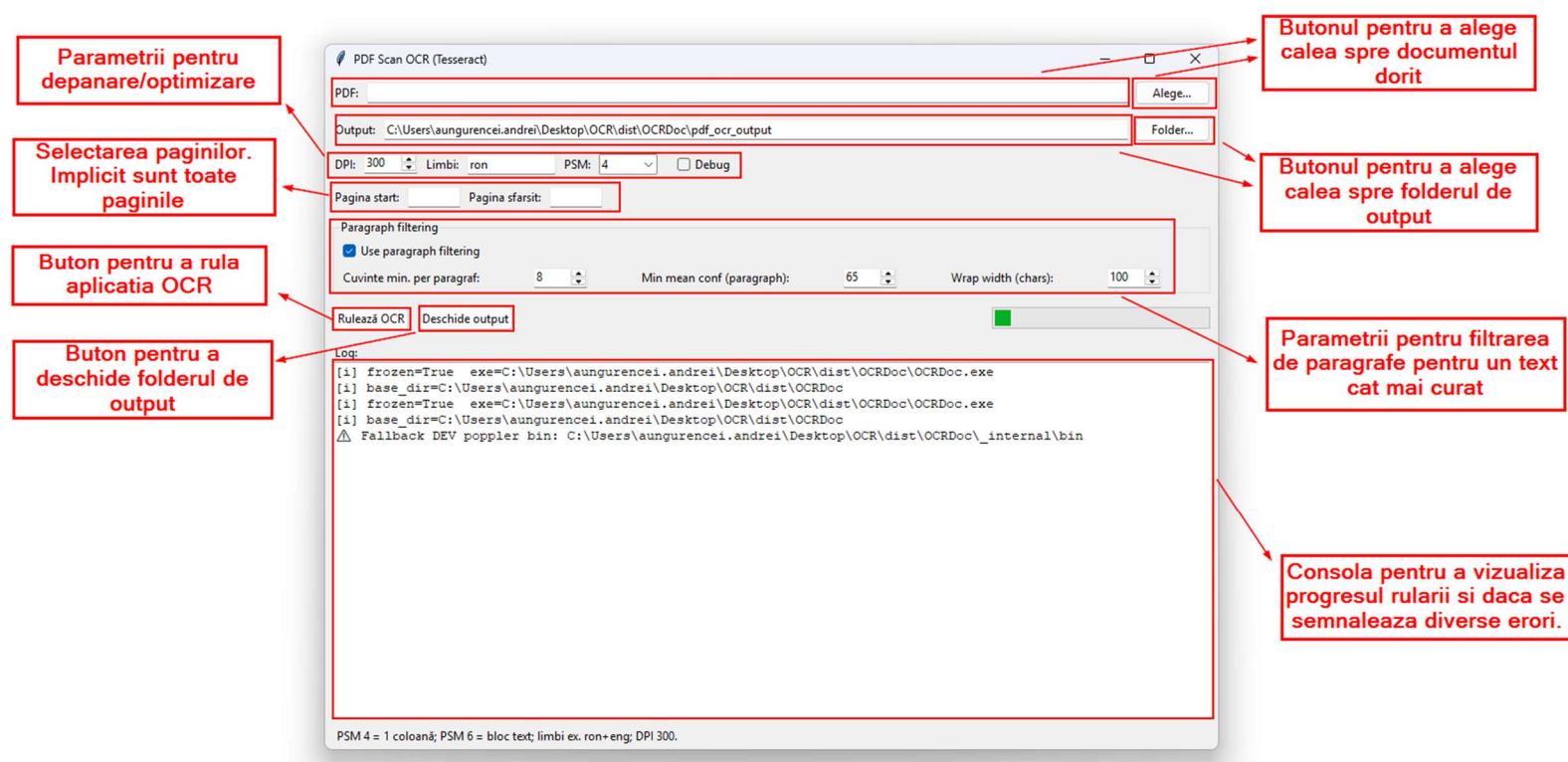
1. Descriere

Documentele PDF care au fost scanate sunt de fapt niste poze, din acest motiv textul scanat nu se poate refolosi decat daca este rescris de mana. Aceasta aplicatie usureaza munca omului de a refolosi acel text intr-un alt fisier de ex. Word. Momentan la versiunea 0.1.0 pe care o lansam aceasta va extrage textul din acel document scanat si le va adauga intr-un fisier text. In urma testelor de catre utilizatori se vor adauga imbunatatiri si la cerere alte functionalitati ale aplicatiei.

2. Instructiuni de utilizare

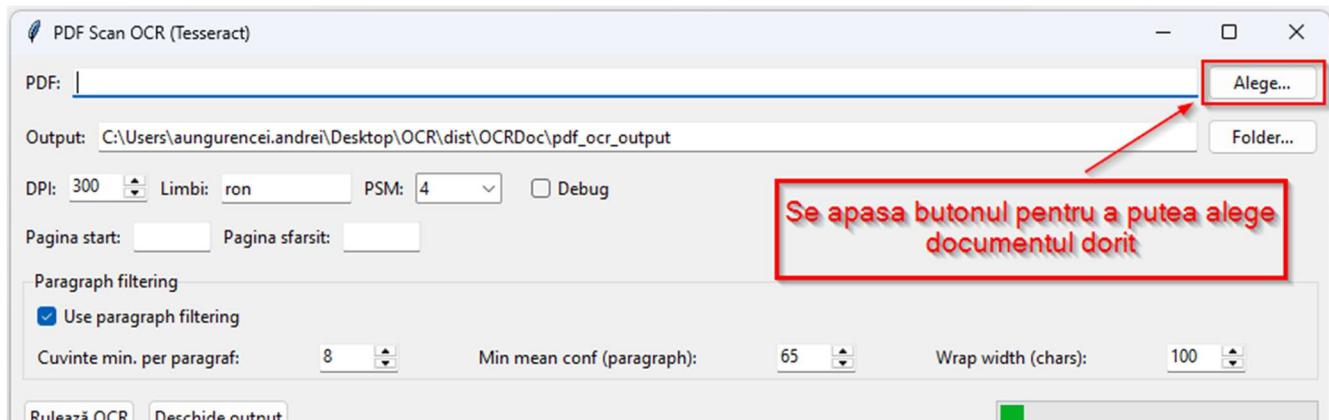
Aplicatia la prima deschidere:



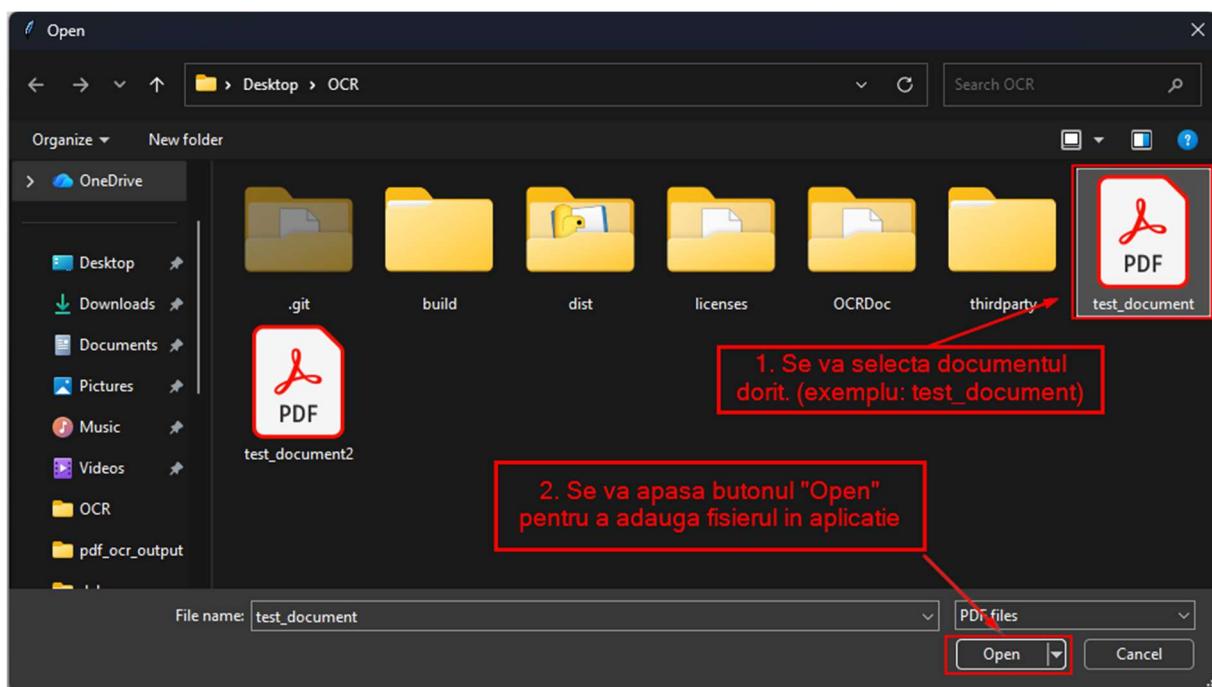


Rularea Aplicatiei

- Pentru a incepe lucru trebuie introdusa calea spre document. Acest lucru se face prin apasarea butonului "Alege...".



- Se va deschide o fereastra unde apoi utilizatorul isi va selecta documentul dorit. Ex:

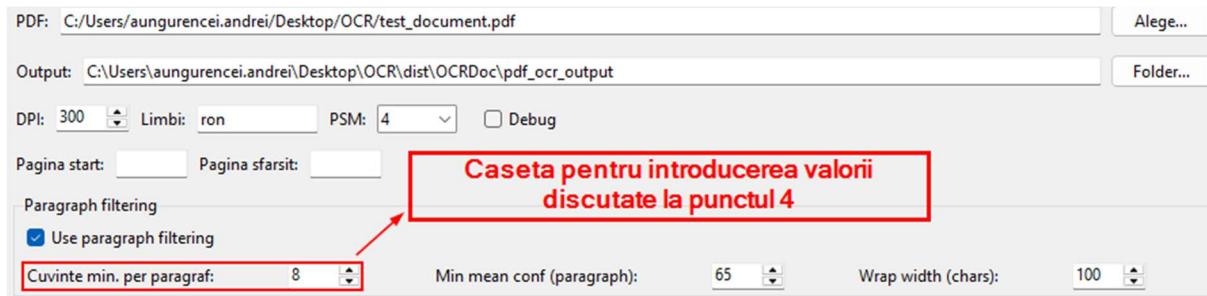


- Optional:** daca se doreste extragerea de text de pe anumite pagini introduceti numarul paginii de start si de sfarsit in casetele cu acelasi nume (ex. 3 si 7 pentru paginile intre 3 si 7; 4 si 4 pentru a extrage doar pagina 4)

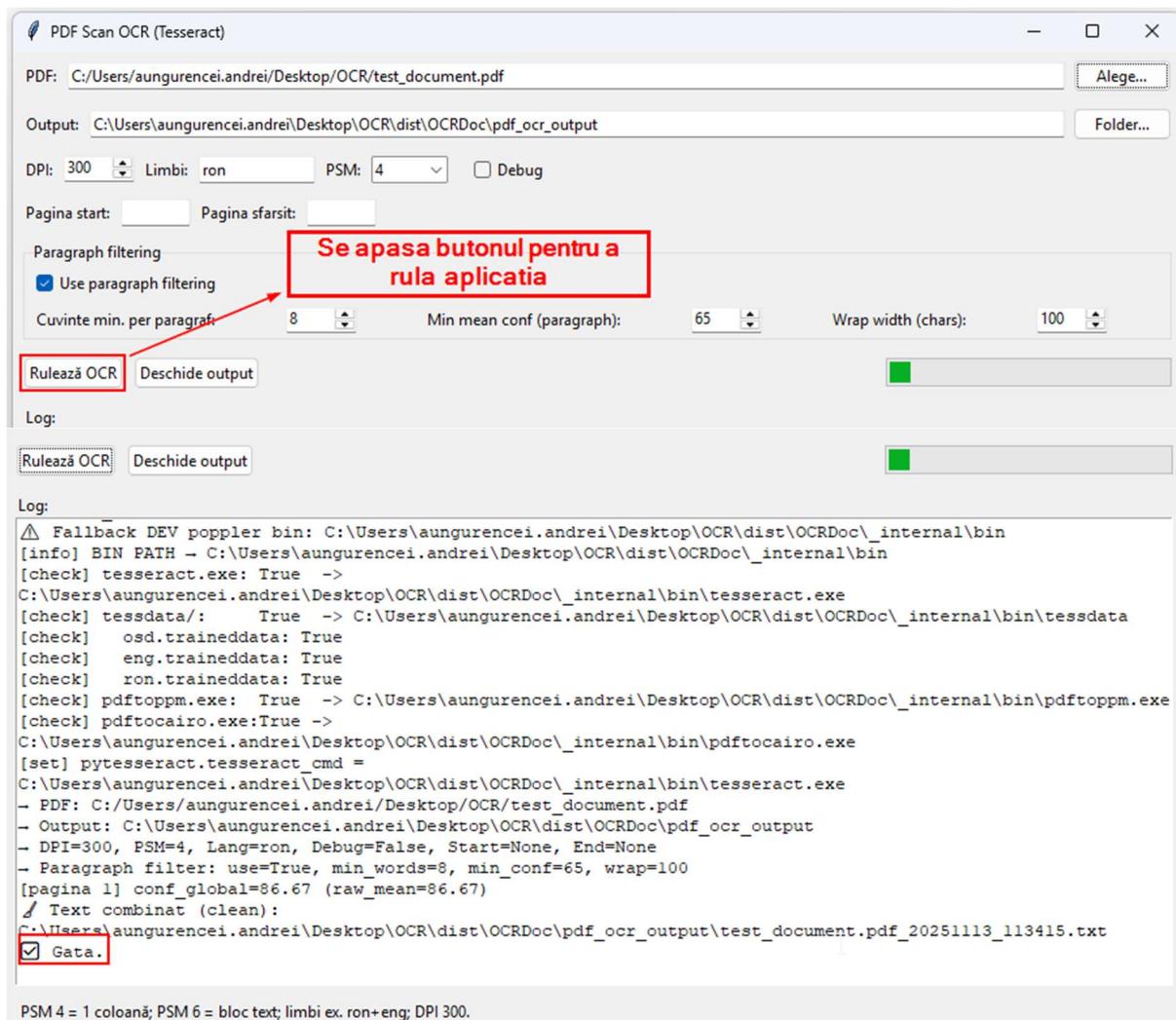
Atentie!!! Pentru extragerea intregului document se va lasa gol deoarece valoarea este implicita.



4. **Optional:** Pentru ca textul rezultat sa fie curat fara alte detectii neconforme este adaugat acest parametru de cuvinte minimum per paragraf care valoarea lui implicita este 8. Aceasta valoare se poate modifica de catre utilizator dupa plac. Se recomanda a se folosi o valoare intre 3 – 8 pentru un text cat se poate de curat. Cei alți parametri nu trebuie modificati.

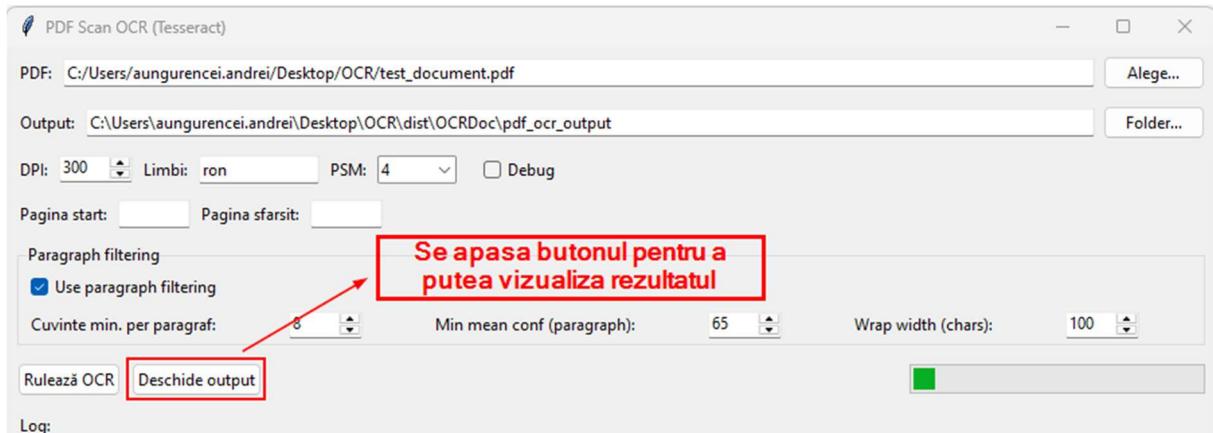


5. Pentru rularea programului se va apăsa butonul "Rulează OCR". Se așteaptă apoi ca în consola să apară "gata"

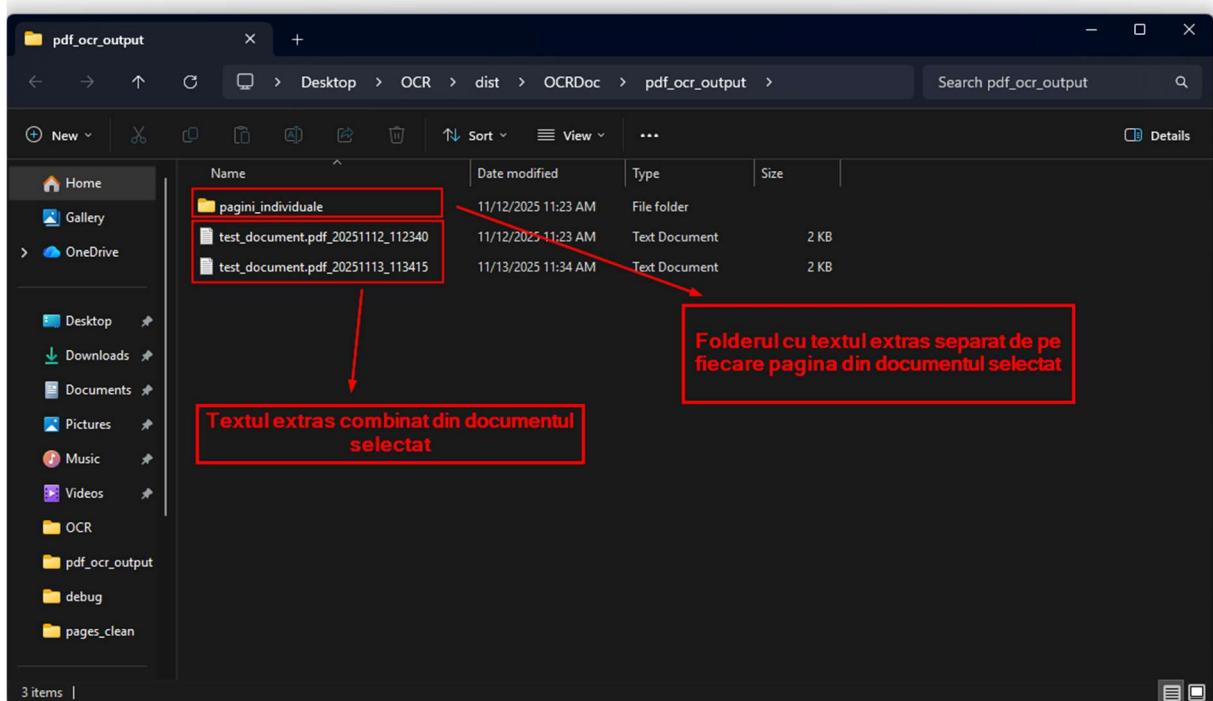


6. Pentru vizualizarea textului extras se va apăsa pe butonul "Deschide output". Se va deschide o fereastră unde veți avea un fișier text cu numele documentului ales și data rularii aplicatiei. În acesta se va găsi textul din toate paginile documentului combinate.

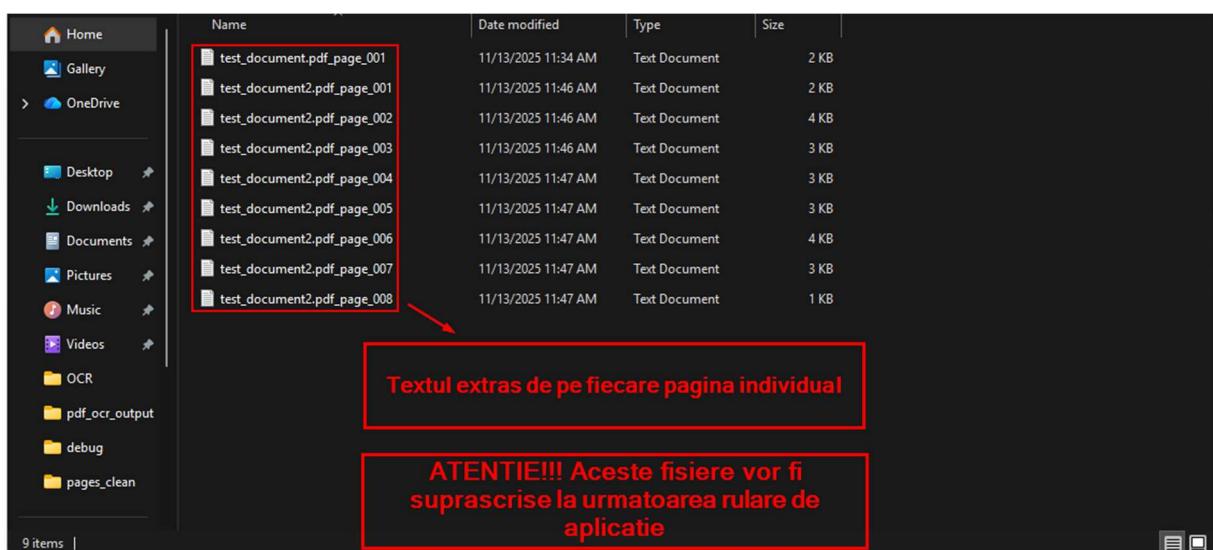
Pe langa acesta va mai exista un folder cu textul pe pagini separate. Acesta folder se va suprascrie de fiecare data cand aplicatia va rula.



Se apasa butonul pentru a putea vizualiza rezultatul



Se va deschide aceasta fereastra unde puteti gasi textul rezultat din OCR



Textul extras combinat din documentul selectat

Folderul cu textul extras separat de pe fiecare pagina din documentul selectat

Textul extras de pe fiecare pagina individual

ATENTIE!!! Aceste fisiere vor fi suprascrise la urmatoarea rulare de aplicatie

