

A Comprehensive Pipeline To predict Stem Cell Differentiation

INTRODUCING A NOVEL PIPELINE LEVERAGING DEEP LEARNING AND GENOMIC DATA TO PREDICT STEM CELL DIFFERENTIATION AND ASSESS CELL HEALTH, ENHANCING REGENERATIVE MEDICINE APPLICATIONS.

1.

MOTIVATIONS

ACCURATE PREDICTION OF STEM CELL DIFFERENTIATION IS CRUCIAL FOR REGENERATIVE MEDICINE BUT REMAINS CHALLENGING WITH CURRENT METHODS. MISMANAGED DIFFERENTIATION CAN LEAD TO TUMORIGENESIS AND CARCINOGENESIS COMPLICATING TREATMENT OUTCOMES.

QUESTIONS

- HOW ACCURATELY CAN OUR PIPELINE PREDICT STEM CELL DIFFERENTIATION USING MACHINE LEARNING AND GENOMIC DATA?
- CAN OUR APPROACH RELIABLY ASSESS THE HEALTH OF STEM CELLS TO PREVENT TUMORIGENESIS?
- WHAT ARE THE KEY GENOMIC MARKERS THAT INFLUENCE STEM CELL DIFFERENTIATION OUTCOMES?
- HOW DOES OUR PREDICTIVE PIPELINE IMPROVE UPON EXISTING METHODS IN TERMS OF PRECISION AND SCALABILITY?

Image 1. An exemple of a tumorigenesis caused by misplaced stem cells.

2.

INTRODUCTION

IN REGENERATIVE MEDICINE, UNDERSTANDING AND CONTROLLING STEM CELL DIFFERENTIATION IS ESSENTIAL DUE TO THEIR POTENTIAL IN TREATING VARIOUS DISEASES AND INJURIES. THIS PROJECT AIMS TO CREATE A MODEL ARCHITECTURE USING COMPUTER VISION AND BIOINFORMATICS TO ANALYZE AND PREDICT STEM CELL COLONY DIFFERENTIATION.

OBJECTIVE

This project aims to develop a model to predict stem cell colony differentiation by segmenting colonies, analyzing DNA/RNA sequences, predicting proteins, and creating an open-source database for further research.

METHODOLOGY

1. Data Collection
2. Segmentation Model Development
3. Differentiation Model Development
4. Protein Prediction
5. Database Creation
6. Integration and Testing
7. Deployment and Maintenance

Image 2. Stem cells Colony under the microscope

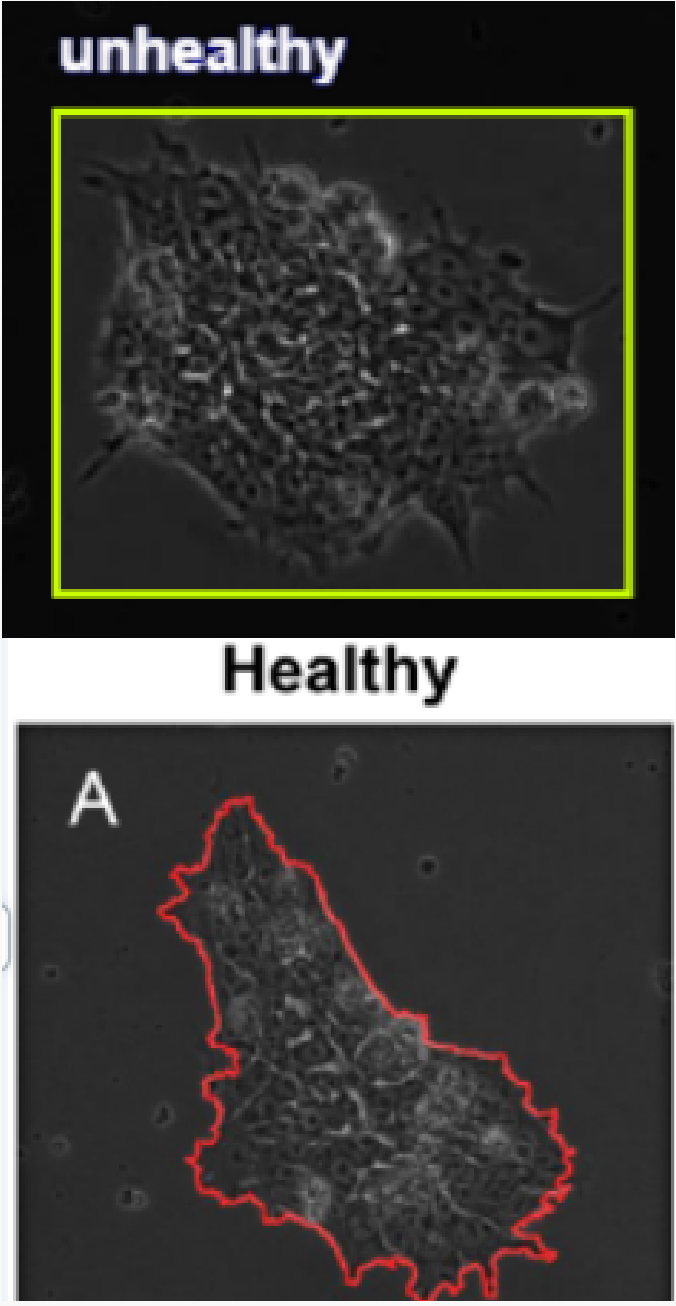
3.

STEM CELLS HEALTH PREDICTION

TO ASSESS STEM CELL HEALTH, WE EMPLOYED TWO MODELS:

- FEATURE EXTRACTION + SVM MODEL: THIS MODEL EXTRACTS RELEVANT FEATURES FROM THE GENOMIC DATA AND USES A SUPPORT VECTOR MACHINE (SVM) TO PREDICT CELL HEALTH.
- YOLOV10 CLASSIFICATION MODEL: THIS MODEL USES THE YOLOV10 ARCHITECTURE TO CLASSIFY CELLS AS HEALTHY OR NOT BASED ON THEIR CHARACTERISTICS.

Image 3. Here is a brief description of this image including sources.



4.

HEALTHY STEM CELLS PREDICTION FEATURES

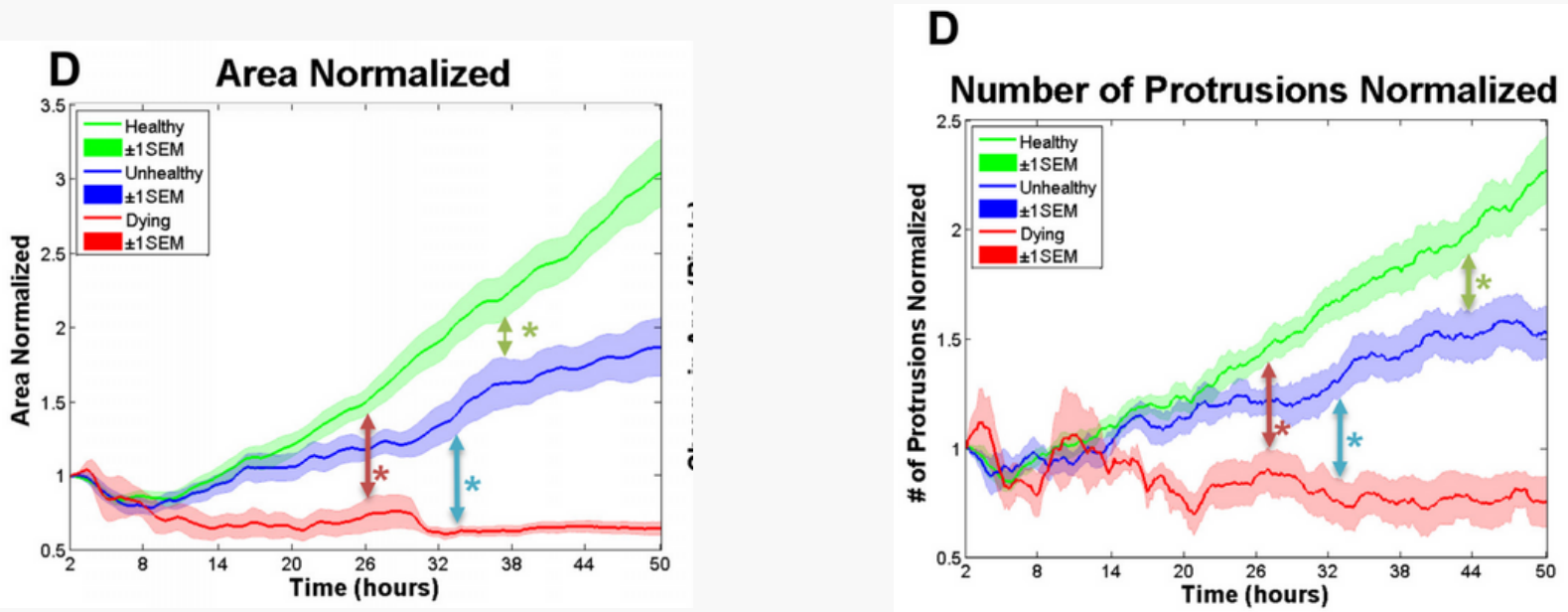
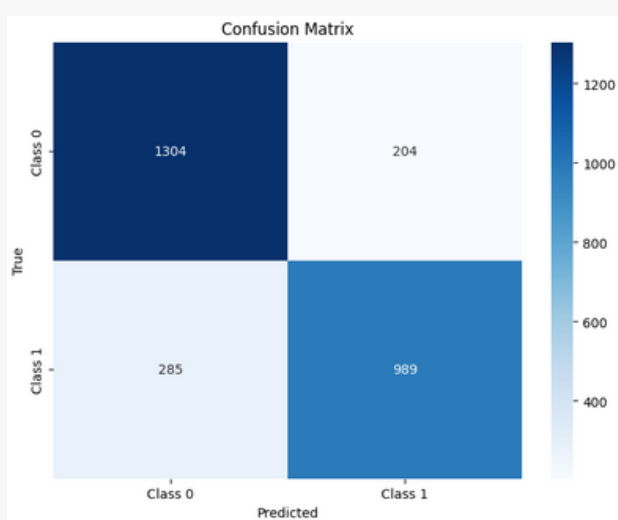


Image 4. Comparison between various Features from the stem cell colonies

5.

HEALTHY STEM CELLS PREDICTION RESULTS

SVM MODEL :



ACCURACY : 0.82422
PRECISION : 0.82900
RECALL : 0.77629
F1 SCORE : 0.8017

YOLOV10 MODEL :



ACCURACY : 0.80
PRECISION : 0.7545
RECALL : 0.7929
F1 SCORE : 0.8222

Image 5. Progress of the loss value across the models

RESEARCH / FINDINGS

Our research identified key proteins serving as biomarkers to predict stem cell differentiation accurately.

- The recent research papers use fluorecence to bind to specific proteins hence identify them
- The Next part summarize some of the findings we have encputered

FINDING 1.

Some proteins like ROR2 indicate that a cell will differentiate to a Neuronal cell. GATA4 protein indicates that a cell will become a cardiac muscle cell

FINDING 2.

NEw Deep Learning techniques like RAG systems that use specific semantic search binded with LLM can interpret and find specific RNA sequences with the stem cell's genetic code

PROPOSED ARCHITECTURE

Our pipeline begins with sequencing the RNA of stem cells. Using a RAG-like system, we identify proteins that correlate with specific differentiation pathways:

- RNA Sequencing: Extract and sequence RNA from stem cells.
- Protein Search: Use a RAG-like system to scan the RNA sequence for known protein biomarkers.
- Biomarker Identification: Identify proteins such as GATA4, which indicates potential differentiation into cardiac muscle cells, among other biomarkers for various cell types.

This process is key to predicting stem cell differentiation and enhancing regenerative medicine applications.

6.

DATASETS

The results of identifying RNA seq of proteins with the RNA sequence of the stem has given mixed results and further analysis and approaches should be investigate. This approach is showinf promise as it accelarates the process of identification

Image 6. A 3D structure of the protein

CONCLUSION

The results of identifying RNA seq of proteins with the RNA sequence of the stem has given mixed results and further analysis and approaches should be investigated. This approach is showing promise as it accelarates the process of identification of cell differentitaion hence making

REFERENCES

European Molecular Biology Laboratory (2023), ENA, <https://www.ebi.ac.uk/ena/browser/view/ERR914288>

PLOS ONE(2016), Evaluating Cell Processes, Quality and Biomarkers in Pluripotent Stem Cells Using Video Bioinformatics, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0148642>