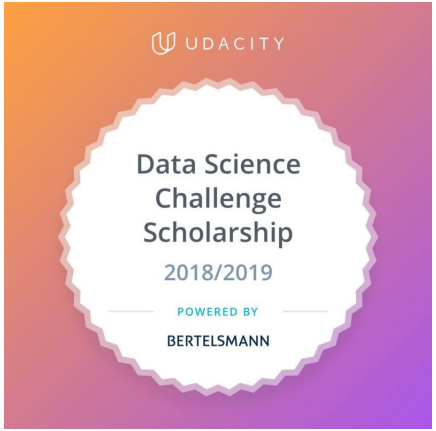
Classification of Genetic Risk for Pediatric Acute Myeloid Leukemia



Health Hackers

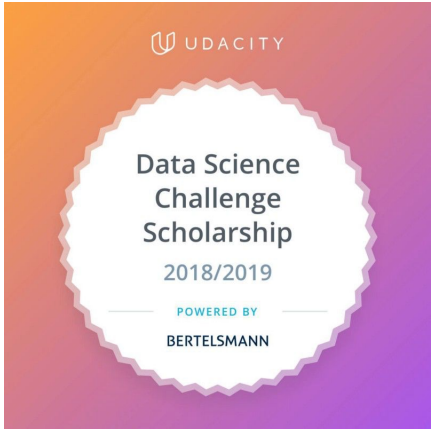
Mahmoud Hamza

Minerva Panda

Úrsula Pérez

Hirza Pimentel

Marwa Qabeel



Introduction

- **MOTIVATION**

Knowing the **genetic risk level** for **pediatric acute myeloid leukemia** (AML) will allow patients to receive the **most effective treatment**.

- **AIM**

To **classify the level of genetic risk for pediatric AML** according to medical **established criteria** and explore the **attributes** more **responsible for each risk level**.

- **HYPOTHESES** for risk level of pediatric AML

❖ Being older is not associated with a higher risk level, it is genetics-related.

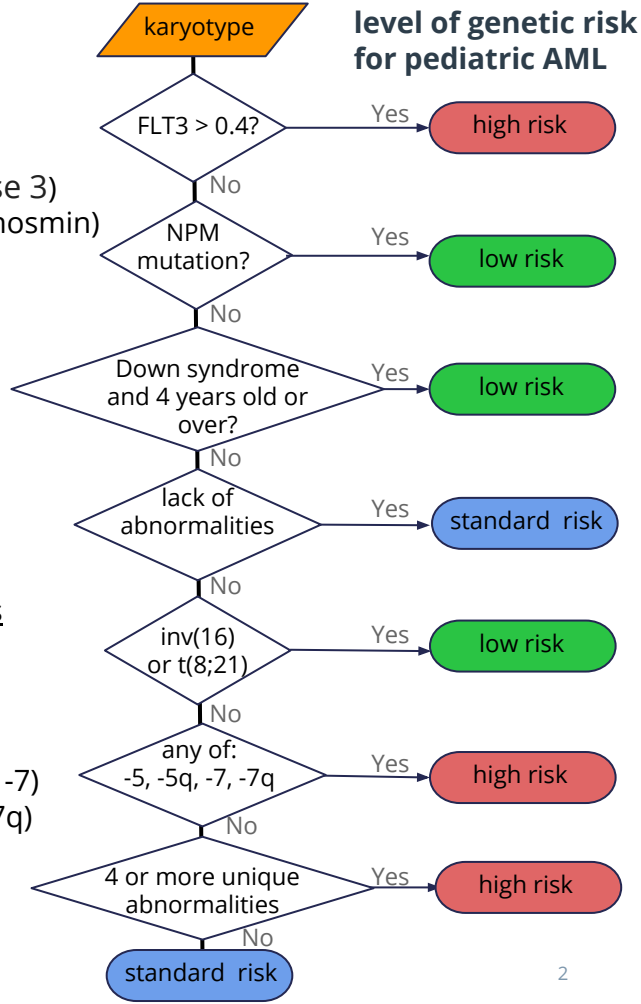
❖ Translocation t(8;21) is the most frequent cause of low risk.

❖ FLT3 mutation is the most responsible cause for high risk (age > 10 years old in the references).

Genes
FLT3 (Fms-like tyrosine kinase 3)
NPM (nucleophosmin)

DNA fragments
inv: inversion
t: translocation

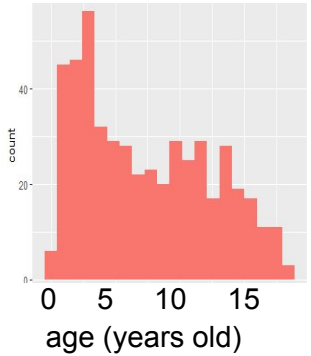
chromosomes
monosomy (-5, -7)
deletion (-5q, -7q)

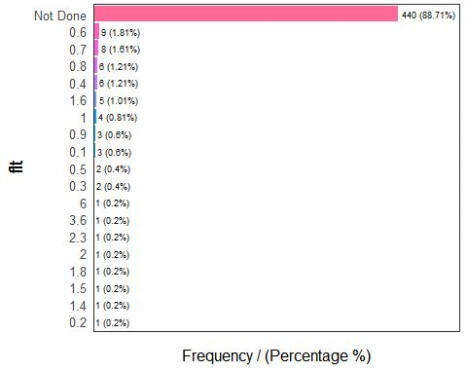


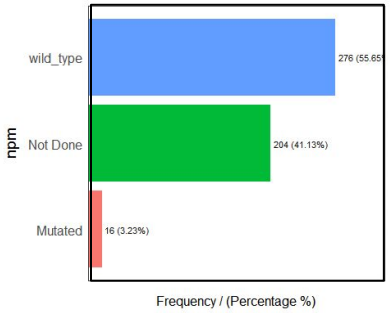
Material & Methods

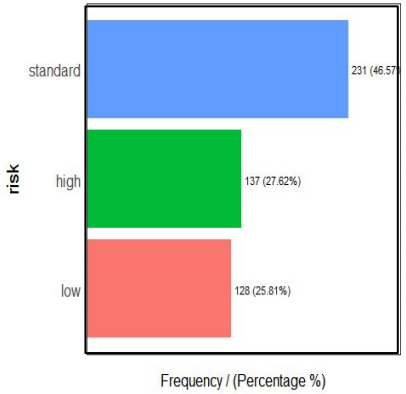
Sample characteristics

504 patients from a hospital in Egypt.









Methods

A) Data cleaning and extracting useful features from data for analyses

B) Classification algorithm for assigning risk scores for pediatric AML to patients, based on the standard medical regulations.

```python
def analyze_data(row_data):

  if row_data['flt3_allelic_ratio'] != 'Not Done' and float(row_data['flt3_allelic_ratio']) > 0.4:
    row_data['riskcause']= 'flt3'
    return 'high'
  elif row_data['npm'] == 'Mutated':
    row_data['riskcause']= 'npm'
    return 'low'
  elif row_data['down_syndrome'] == 1 and int(row_data['age']) >= 4:
    row_data['riskcause']= 'down_syndrome'
    return 'low'
  elif row_data['abnormalities'] == '':
    return 'standard'
  else:
    #row_data['riskcause']= 'abnormality in karyotype'
    for sub_idx in range(len(row_data['abnormalities'])):
      abnormality = row_data['abnormalities'][sub_idx]
      if 'inv(16)' in abnormality:
        row_data['riskcause']= 'inv(16)'
        return 'low'
      elif 't(8;21)' in abnormality:
        row_data['riskcause']= 't(8;21)'
        return 'low'
      elif abnormality == '-7q':
        row_data['riskcause']= '-7q'
        return 'high'
      elif abnormality == '-7':
        row_data['riskcause']= '-7'
        return 'high'
      elif abnormality == '-5q':
        row_data['riskcause']= '-5q'
        return 'high'
      elif abnormality == '-5':
        row_data['riskcause']= '-5'
        return 'high'
      elif len(set(row_data['abnormalities'])) >= 4:
        row_data['riskcause']= 'four abnormalities and higher'
        return 'high'
  row_data['riskcause']= 'not low not high'
  return 'standard'
```

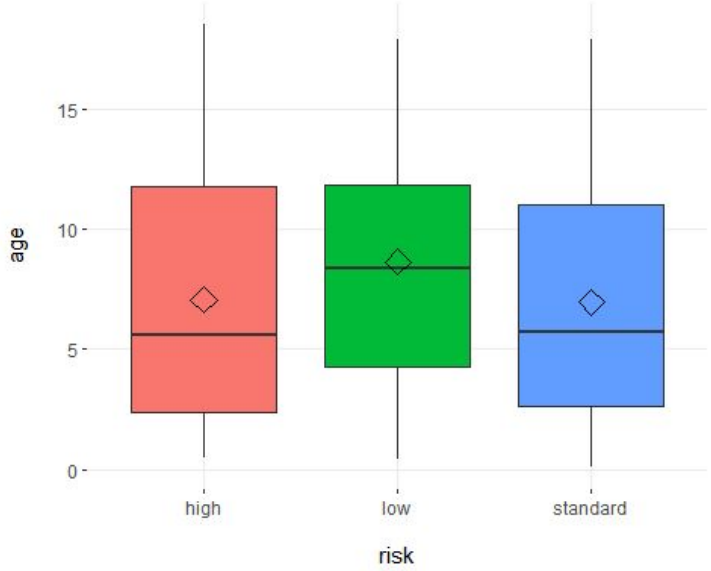
Code link: https://goo.gl/643Qhq

C) Statistical analyses (descriptive statistics and Kendall's Tau correlation)

Results & Conclusions

a) age vs. risk level

Kendall's Tau correlation = -0.032 (*p*-value = 0.36)

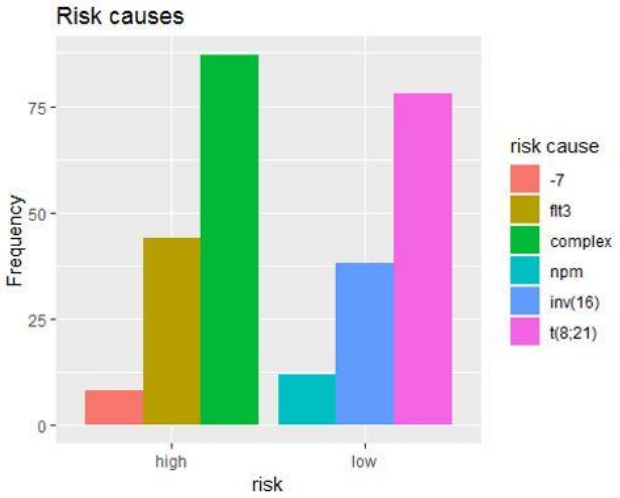


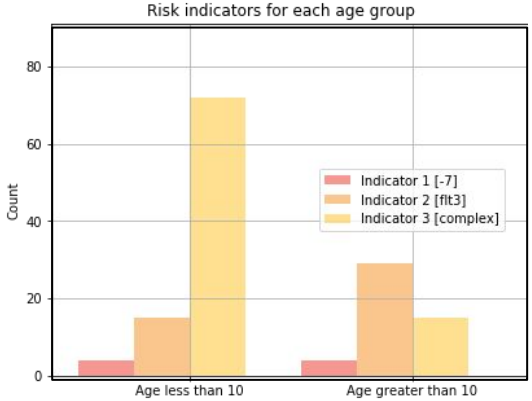
- Age is not statistically associated with risk level (*p*-value = 0.36), with a Kendall's Tau correlation value close to zero (-0.032). Therefore, older and younger infants are almost equally predisposed to a low/standard/high risk level of pediatric AML.
- This is in accordance with our hypothesis: the risk level is genetics-related but not age-related.

All the patients were correctly classified as expected, following the medical criteria.

Results & Conclusions

b) risk causes for high and low risk levels



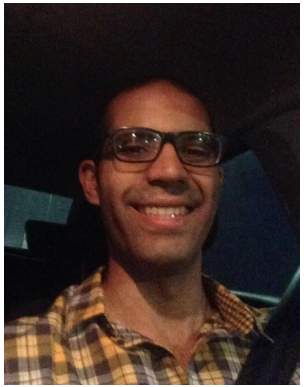


age rank	high: FLT3
(0,10]	15/91
(10,18.5]	29/48

- Four or more unique abnormalities (complex karyotyping) was the most frequent risk cause for the high risk of pediatric AML considering all age ranges. We hypothesized that FLT3 would be the most frequent cause, but this is only true for patients greater than 10 years old.
- For low risk of pediatric AML, the translocation t(8;21) was the most frequent cause. This is in accordance with our hypothesis.

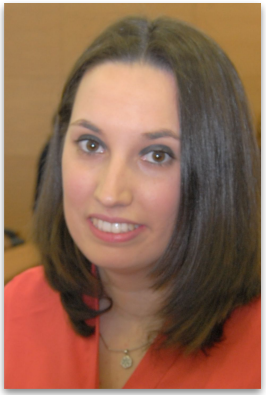
Future work: to predict FLT3 value (lower or greater than 0.4) from the other variables with machine learning techniques. FLT3 is crucial to know the risk level but requires a long analysis in the laboratory.

Health Hackers



Mahmoud Hamza

Minerva Panda



Úrsula Pérez

Hirza Pimentel

Marwa Qabeel

References

- Stack Overflow contributors, Correlation between continuous and categorical variables in Python; 07/20/2018; (https://stackoverflow.com/questions/44694228/how-to-check-for-correlation-among-continuous-and-categorical-variables-in-python)
- Panagiotis D. Kottaridis and others, The presence of FLT3 in patients with (AML); 07/18/2018; (http://www.bloodjournal.org/content/98/6/1752.full?sso-checked=true)
- Kalliopi N. Manola, Cytogenetics of pediatric acute myeloid leukemia; 07/21/2018; (http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0609.2009.01308.x/pdf)
- Children's oncology group, COG AAML0531 for newly diagnosed Acute Myeloid Leukemia in Children; 07/18/2018; (https://www.childrensoncologygroup.org/index.php/aaml0531/)
- Marry van den Heuvel-Eibrink et al., Pediatric AML: From Biology to Clinical Management; 07/18/2018; (http://www.mdpi.com/2077-0383/4/1/127)
- Hamid Bolouri and others, Comprehensive characterization of pediatric AML reveals diverse fusion oncoproteins and age-specific mutational interactions; 07/18/2018;(https://www.biorxiv.org/content/early/2017/10/10/125609)