2020 Annual Meeting of the Population Association of America

Applied Machine Learning for Neonatal Mortality Risk Assessment: A Case Study Using Public Health Data from São Paulo - Brazil

Carlos Ed Beluzo
linkedin.com/in/cbeluzo

Luciana Alves | *Rodrigo Bresan* | *Natália Arruda* | *Ricardo Sovat* | *Tiago Carvalho*

Background

- Computer Science Professor @ *Federal Institute of São Paulo (IFSP)*
  - Data Science postgraduate course creation committee member @ *IFSP*
  - Research group deputy leader @ *PICAp - IFSP*
- Ph.D Candidate @ *University of Campinas, Department of Demography - (Unicamp)*
  Supervisor: Luciana Correia Alves
- Data Science Research Project Manager and co-PI (Bill & Melinda Gates Foundation grant)
  - *Decision-Making Support Platform Based on Visual Analytics and Machine Learning to Subsidize Public Politics Focused on Gestational Health*
- B.A. Informatics (2002), MSc. in Mechanical Engineering (2006), more than 10 years experience with Databases & Software Development
- **Interesting fields**: Applied Computer Systems; Demography methods; Data Science; Project Management; Agile Methodologies; Databases; Big Data.

Agenda

1. Overview
2. The Proposal
3. The Dataset
4. The Method
5. Experiments and Results
6. Conclusion and Research Directions

Overview

- **Infant mortality**
  - Reflection of a complex combination of factors:
    - Biological, socioeconomic, health care, etc
  - Requires various data sources for a thorough analysis;
    - Specialized tools/techniques to deal with a large volume of data.

- **Research Question:**

  Is it possible to "predict" neonatal mortality using this framework?

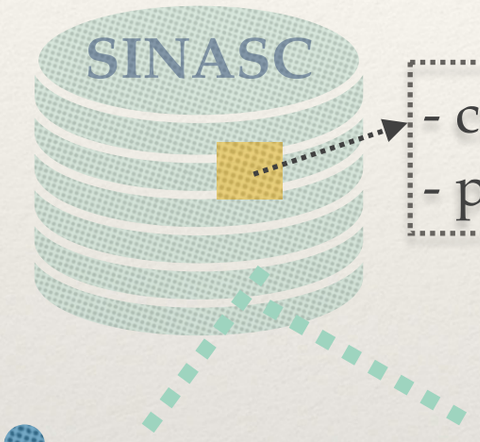
- **Machine Learning** (ML) has been applied to solve problems from many domain
  - Presents great potential for this problem too.

The Proposal

- A method to perform **neonatal death risk assessment** using ML
- Using mother, pregnancy care and child at birth features
- Public health dataset containing neonatal samples (deaths/alive)
- Encodes feature vectors into images and classifies images using ML
  - Custom convolutional neural network (CNN)
  - As results the method classifies samples as death or alive
- **Method is able to detect death samples with accuracy of 90.61%.**

The Dataset

Brazilian Information System of Live Births

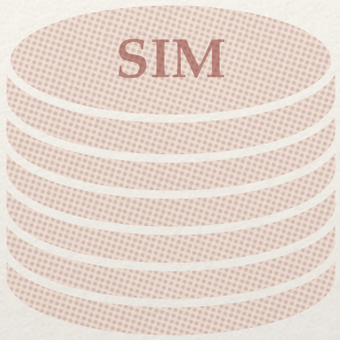


- city = “São Paulo”
- period = [2012, 2018]

birth location
birth date
sex
APGAR score
weeks of gestation
anomalies
. . .

age
schooling
occupation
children
delivery type
prenatal appointments
. . .

Brazilian Information System of Mortality

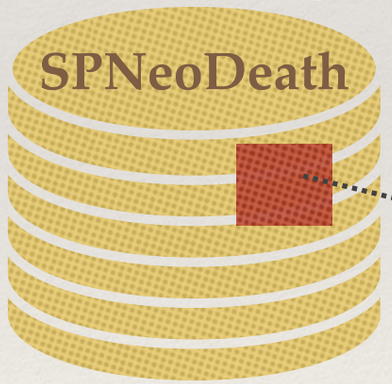


nu_do
birth date

Linkage

nu_do
death date

- 1.4 mi samples
- 0.??% death



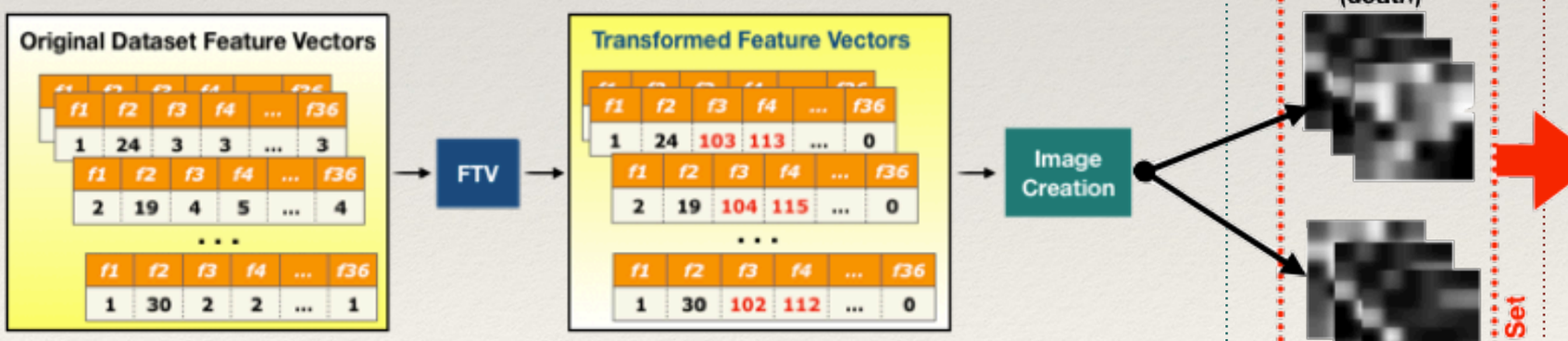
- Neonatal deaths only
- 4 mil samples

WHERE *SIM.nu_do == SINASC.nu_do*
AND *(SIM.death_date - SINASC.birth_date) < 29*

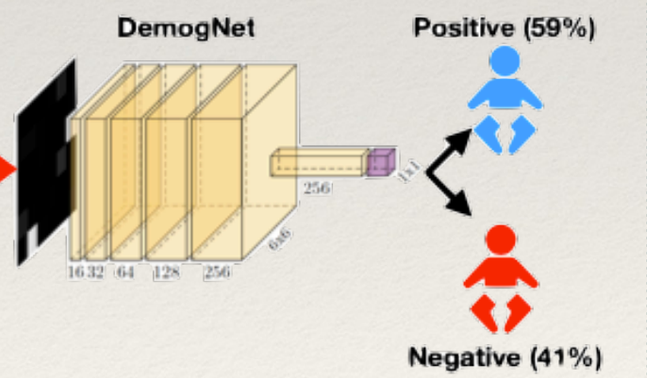
The Method

1 - Features Selection

2 - Feature Transformation



3 - Model Creation



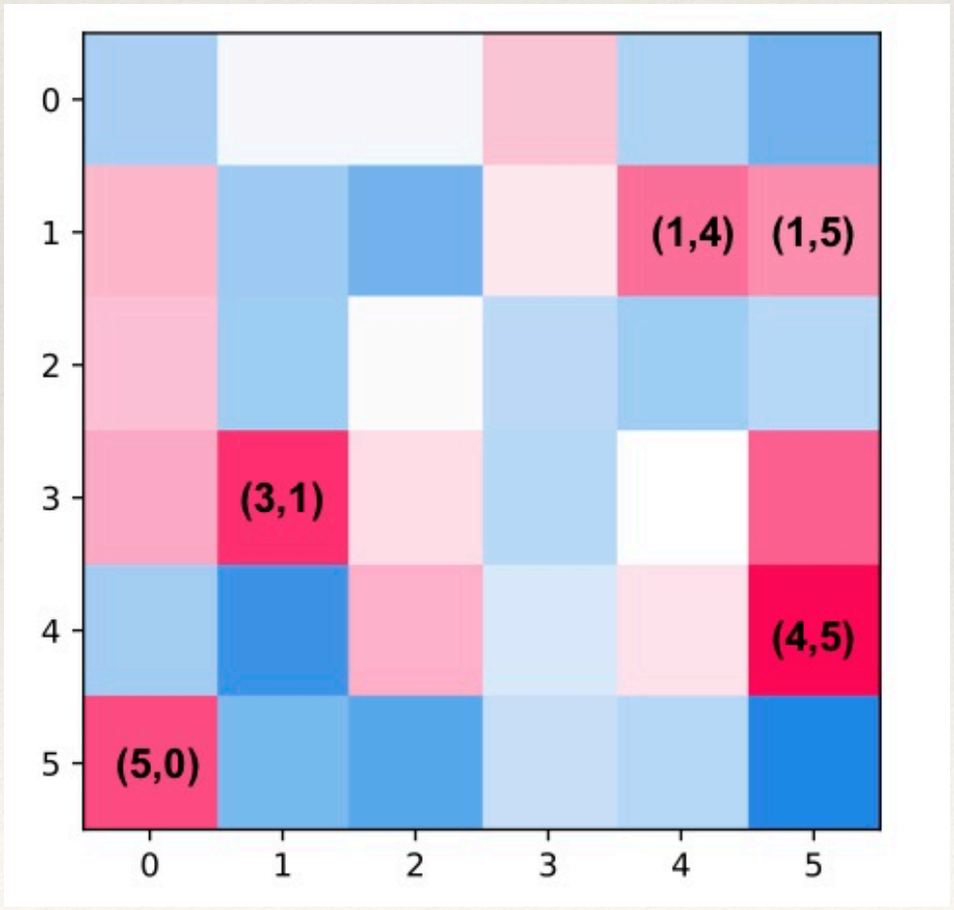
Experiments and Results

- Round #1: **DemogNet with Balanced Dataset**
- Round #2: **DemogNet with Unbalanced Dataset**
- Round #3: **Standard Classifiers for Benchmark**
  - *SVM (Support Vector Machines)*
  - *KNN (K-Nearest Neighbor)*
- Round #4: **Providing Model Understanding**
  - *SHAP (SHapley Additive exPlanation)*

Experiments and Results

Table 2. Comparison against standard classification methods

	ACC. (Balanced)	AUC. (Balanced)	ACC. (Unbalanced)	AUC. (Unbalanced)
DemogNet	**0.89**	**0.95**	**0.91**	**0.96**
KNN	0.83	0.90	0.84	0.90
SVM	0.58	0.91	0.58	0.91



Conclusion and Research Directions

- A new method to address neonatal death risk problem
- SINASC X SIM (1.4m) categorical: mother, pregnancy care, child features at born, etc.
- A new approach to encode this categorical data into small gray scale images.
- Problem modeled as a binary classification of death and living classes
- **DemogNet** implementation, new CNN architecture
- Effectiveness of model classification, achieving an AUC value of 0.96
- Experiments demonstrate that DemogNet outperform standard machine learning methods;
- Experiments to understand what contributes to model final answer,
- **Method limitations:** *Post born features | dataset bias*
- **Future works**: Applies with other datasets (**BRNeoDeath** is in progressing)

Acknowledgment

This work is supported by:

- *Bill & Melinda Gates Foundation [OPP1201970];*
- *Brazilian Ministry of Health, Brazilian National Council for Scientific and Technological Development [443774/2018-8];*
- *NVIDIA corporation for supporting our research with a TITAN Xp GPU donation;*
- *Center for Information Technology Renato Archer (CTI).*

PAA 2020
ANNUAL MEETING
WASHINGTON, DC · APRIL 22-25

Thanks

cbeluzo@gmail.com