CHOC Research
GO BEYOND

Improving Clinical Outcomes for Children Using Data Science

Louis Ehwerhemuepha, PhD

Director, Computational Research

Expectations

- About CHOC
- Mission and Vision of the Research Computational and Data Science (RCS) Unit
- RCS Team, Research Areas, and Research Snapshot
- Example studies
  - Readmissions
  - Juvenile Dermatomyositis
  - Other studies
- Personal Journey
- Conference and Datathon

CHOC at a Glance

2 HOSPITALS

334 BEDS

CHOC Hospital in Orange

54 BEDS

CHOC at Mission Hospital

5 CENTERS OF EXCELLENCE

- Heart
- Research
- Orthopedic
- Neurology / Neurosurgery
- Oncology

ACCOLADES

BEST CHILDREN'S HOSPITALS
U.S.News & WORLD REPORT
RANKED IN 7 SPECIALTIES

THE LEAPFROG GROUP
TOP CHILDREN'S HOSPITAL
2021

CHOC Research
GO BEYOND

CHOC at a Glance

LEADING-EDGE PROGRAMS

A PEDIATRIC MENTAL HEALTH SYSTEM of care poised to serve as a national model

THE THOMPSON AUTISM CENTER brings the latest treatments and resources to help families thrive

ONE OF THE NATION'S ONLY dedicated cancer programs for teens and young adults

ACADEMIC AFFILIATION

UNIVERSITY OF CALIFORNIA, IRVINE

DESIGNATIONS

TWO
pediatric trauma centers:
Level 1
Level II

GOLD LEVEL
extracorporeal life support (ECLS) center

LEVEL 4
NICU

LEVEL 4
pediatric epilepsy center

MAGENT RECOGNITION from the American Nurses Credentialing Center

PROVIDERS

30 More than 30 specialty areas treated, from allergy to urology

700 More than 700 providers on the medical staff

500 More than 500 residents, fellow and medical students who undergo pediatric training at CHOC every year

4,700 Associates

250k Children cared for every year

CHOC Research
GO BEYOND

CHOC Research
GO BEYOND

RESEARCH AND COMPUTATIONAL SCIENCES (RCS)

Terence Sanger, MD, PhD
Chief Scientific Officer

Phuong Dao, JD
Executive Director,
Research Operations

Louis Ehwerhemuepha, PhD
Director, Computational Research

Don Wen, BS

Supervisor,
Data Services

Visiting Scientists: 3

Graduate Interns: 2

Sumiko Abe, PhD
Senior RCS
(Advanced Imaging)

Peyman Kassani, PhD
Senior RCS
(Computer Vision)

Chloe Martin-King, PhD
RCS II
(Computer Vision)

Ricardo Aguilar, MS
RCS I
(Machine Learning)

Tatiana Moreno, BS
RCS I
(Research Data Services)

Aline Rohloff, MS
RCS I
(Generalist)

Ryan Kassab, MS
Grant-Funded RCS

Madhura
Baxi, PhD

RCS II

Quinn Gates, MS
Grant-Funded RCS

Mission

Provide research data, computational, and data science research services at CHOC and conduct independent research towards advancement of pediatric medicine.

Vision

To be a leading national destination for computational and data science research in pediatrics.

Research areas

- Retrospective analyses of EHR data. CHOC database vs Cerner(multicenter deidentified)
  - What happened?
  - Why did it happen?
  - How can we prevent it from happening again if undesirable?
- Machine learning prediction models on structured/tabular EHR data.
- Deep learning models and related artificial intelligence models on unstructured data (notes, images, etc.).
- Empirical analyses and computational intelligence **including reinforcement learning**.
- Deployment of statistical, machine learning, and AI models.
  - Build computational systems to deploy real time clinical models
  - Integrate with EHR workflow.

Since Inception

Currently deployed models and models to be deployed

- Hospital readmission model | Reduced readmission rates | $1.0 million per year of overall savings (Collaboration with Dr. Lusk, Dr. Weiss, Dr. Pugh, etc.)
- Rising risk – Refitting for population health (with Dr. Weiss, Dr. Golden)
- Clinic no show – final phases of model development prior to clinical deployment (Dr. Weiss, Dr. Golden)
- Autism triage model
- Multiple other clinical administration tools powered by LLMs and other NLP models deployed by our partners

More than 250 research studies

- Research services (via intake portal): 195+
- Independent research in collaboration with clinicians: 35+
- Publications: 45-50 (More than 1 publication per month since inception not including studies published with support from Research Data Services).

Readmissions Version 2.0

CHOC Research
GO BEYOND

Why update model?

- Impact of interventions can plateau based on weight of existing risk factors
- New risk factors may arise
- No model is perfect, so there is always room for improvement
- SARS-CoV-2 and the pandemic changed utilization
- Exploit more recent patterns to build model for interventions
- Adopt post-deployment recommendations from clinical team

Which new variables were considered?

- Planned admission status
- Pediatric comorbidity index
- 36 diagnosis categories (up from 18) including history of prematurity
- Home devices (new orders and existing) such as feeding tubes
- Intensive care history

Modeling

- Logistic regression remains the most parsimonious model with only a slight drop in performance against other (non-parametric) machine learning algorithm including deep learning.
- Adopted logistic regression here again
- Data for training and test: January 2016 to 2022
- Metrics for deployment: January 2023 to August 2023

Findings and risk factors

Variable		Odds ratio	p value
Planned admissions		0.7 (0.64, 0.77)	< 0.001
Previous ED visits in 6 months (Ref: 0)			
	1	1.53 (1.41, 1.66)	< 0.001
	2	1.78 (1.59, 1.99)	
	3 or more	2.38 (2.13, 2.66)	
Previous unplanned readmission, 6 mos (Ref: 0)			
	1	1.47 (1.32, 1.64)	< 0.001
	2	1.89 (1.60, 2.23)	
	3 or more	3.15 (2.63, 3.77)	
Previous hospitalizations, 6 mos (Ref: 0)			
	1	1.64 (1.49, 1.80)	< 0.001
	2	1.98 (1.73, 2.26)	
	3 or more	1.92 (1.64, 2.24)	
Length of stay (Ref: < 2 days)			
	2 - 3 days	1.23 (1.14, 1.34)	< 0.001
	4 - 6 days	1.57 (1.43, 1.73)	
	7 days or more	2.07 (1.88, 2.28)	

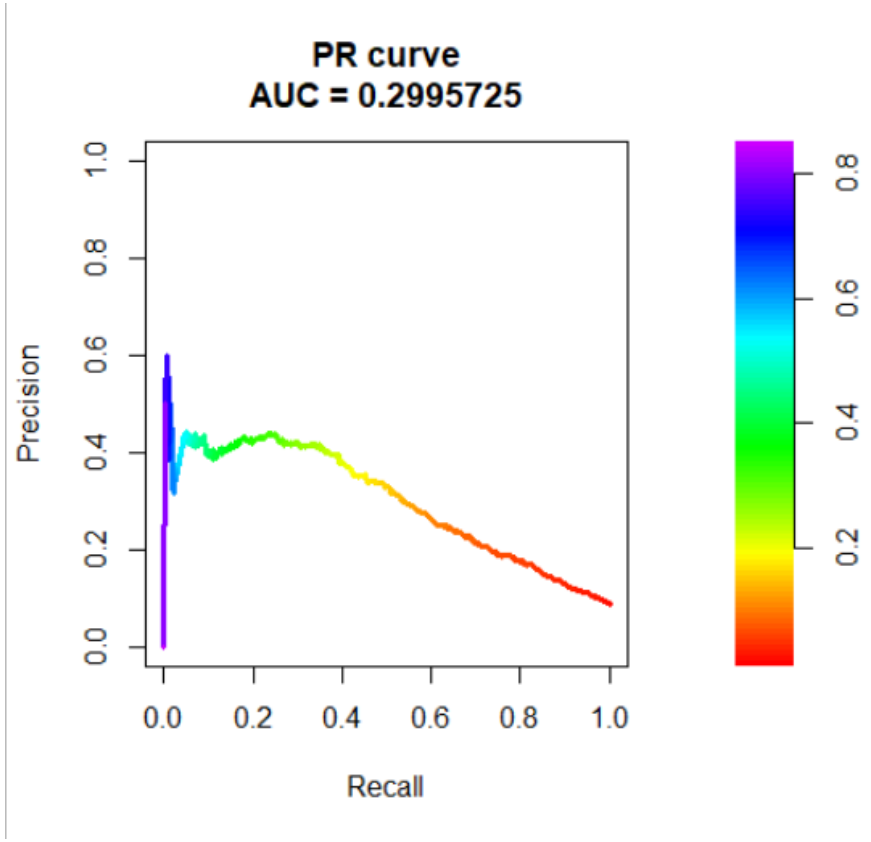
Condition	Odds ratio	p value
Malignant neoplasms	3.27 (2.96, 3.62)	< 0.001
Hypertensive disorders (I10-I16)	1.41 (1.22, 1.62)	< 0.001
Hematological conditions	1.31 (1.23, 1.42)	< 0.001
History of Prematurity	1.24 (1.10, 1.40)	
Neurological conditions	1.20 (1.12, 1.30)	< 0.001
Birth disorders (Q00-Q89)	1.18 (1.09, 1.28)	< 0.001
Genitourinary disorders	1.17 (1.07, 1.29)	< 0.001
Mental/behavioral conditions (F01-F48)	1.17 (1.06, 1.29)	0.003
Chromosomal abnormalities (Q90-Q99)	1.16 (1.02, 1.31)	0.027
Musculoskeletal/connective diseases	1.12 (1.03, 1.23)	0.012
Metabolic diseases	1.12 (1.04, 1.21)	0.004
Endocrinologic condition	0.90 (0.81, 1.01)	0.069
Skin/subcutaneous diseases	0.87 (0.79, 0.95)	0.004
Viral infections	0.84 (0.76, 0.93)	0.001
Bacterial infections	0.81 (0.73, 0.89)	< 0.001
Influenza	0.78 (0.59, 1.00)	0.056
Physical injuries/trauma (S00-S99, T14)	0.73 (0.63, 0.85)	< 0.001

Home devices

Variable	Odds ratio (95% CI)	p value
G-Tubes and other feeding tubes	1.24 (1.13, 1.36)	< 0.001
~~Orthopedic devices~~	~~1.10 (0.98, 1.22)~~	~~0.092~~
Home oxygen device	1.22 (1.10, 1.35)	< 0.001
Ventricular assistive devices (VAD)	1.54 (1.36, 1.74)	< 0.001

Simulated live performance metrics

- AUROC: 0.80 (0.77, 0.82)
- Balanced metrics
  - Sensitivity: 73%
  - Specificity: 73%
  - PPV: 21%
  - NPV: 97%
  - NNE: 4.8



How should we operationalize the model?

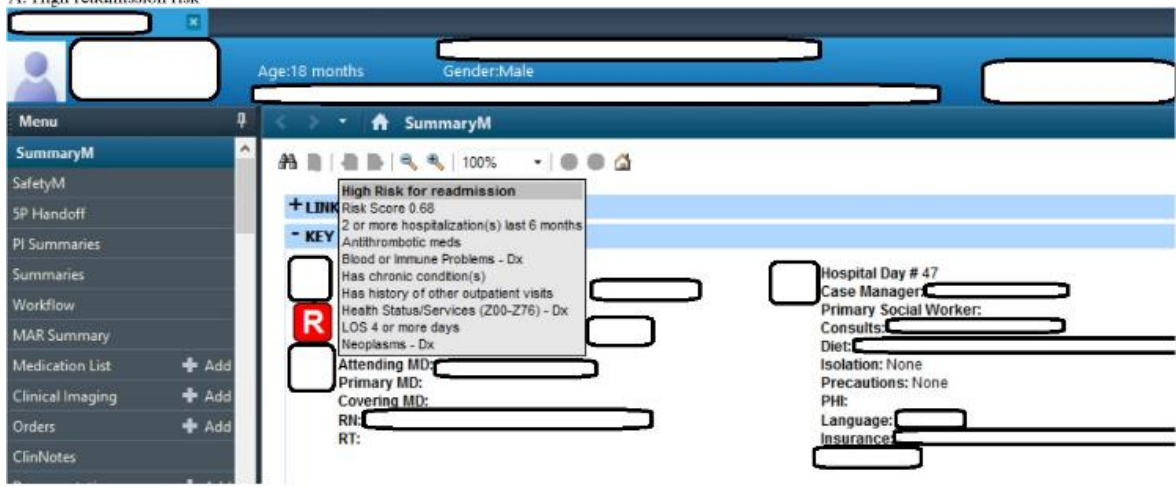
Risk, Specificity	Sensitivity	Positive predictive value	Negative predictive value	F1 Score	NNE	Predicted probability threshold
High risk, 95	37%	40.4	94	0.38	2 of 5	0.23
Moderate risk, 99	51%	32.3	95	0.40	1 of 3	0.15

Workflow and adoption

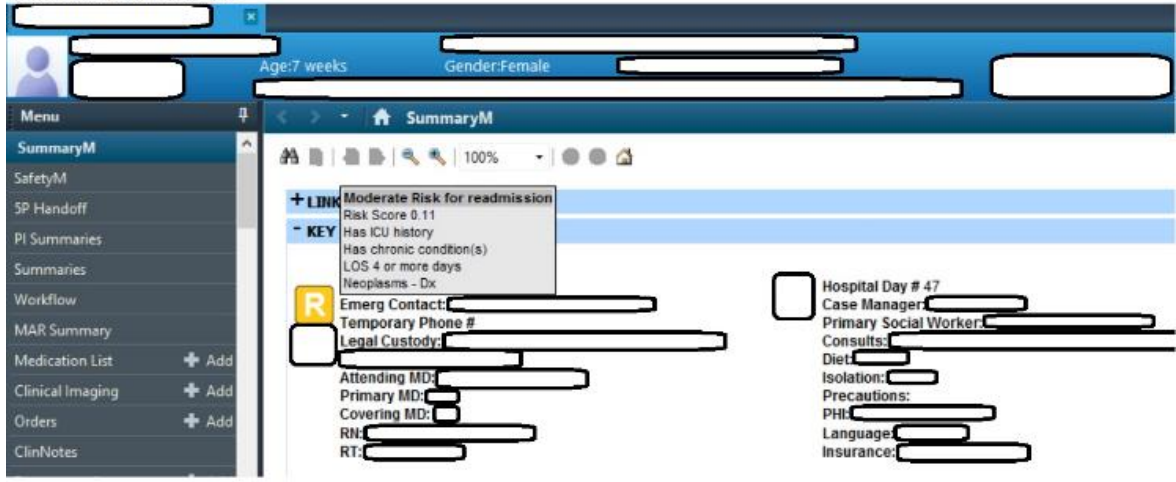
- Icon on SummaryM page
- Risk stratified
- Lists variables contributing to score

- Increased awareness by all
- Trigger for discussion on rounds

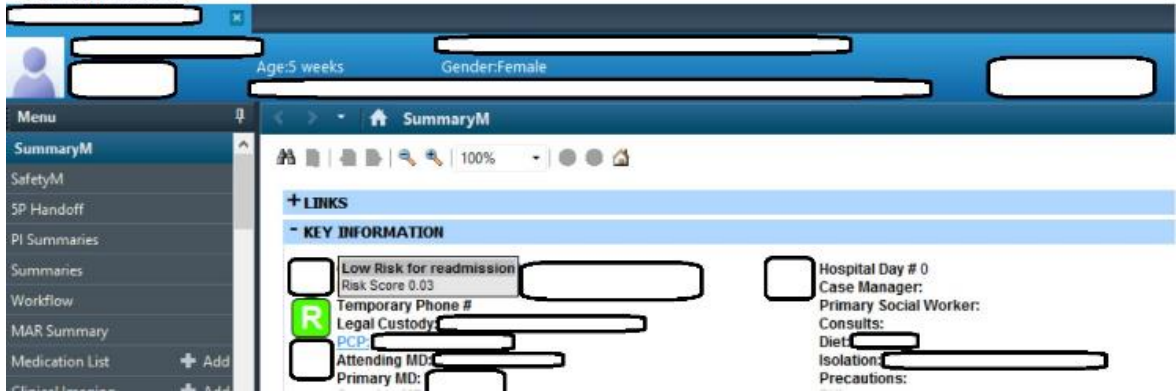
A. High readmission risk



B. Moderate readmission risk



C. Low readmission risk



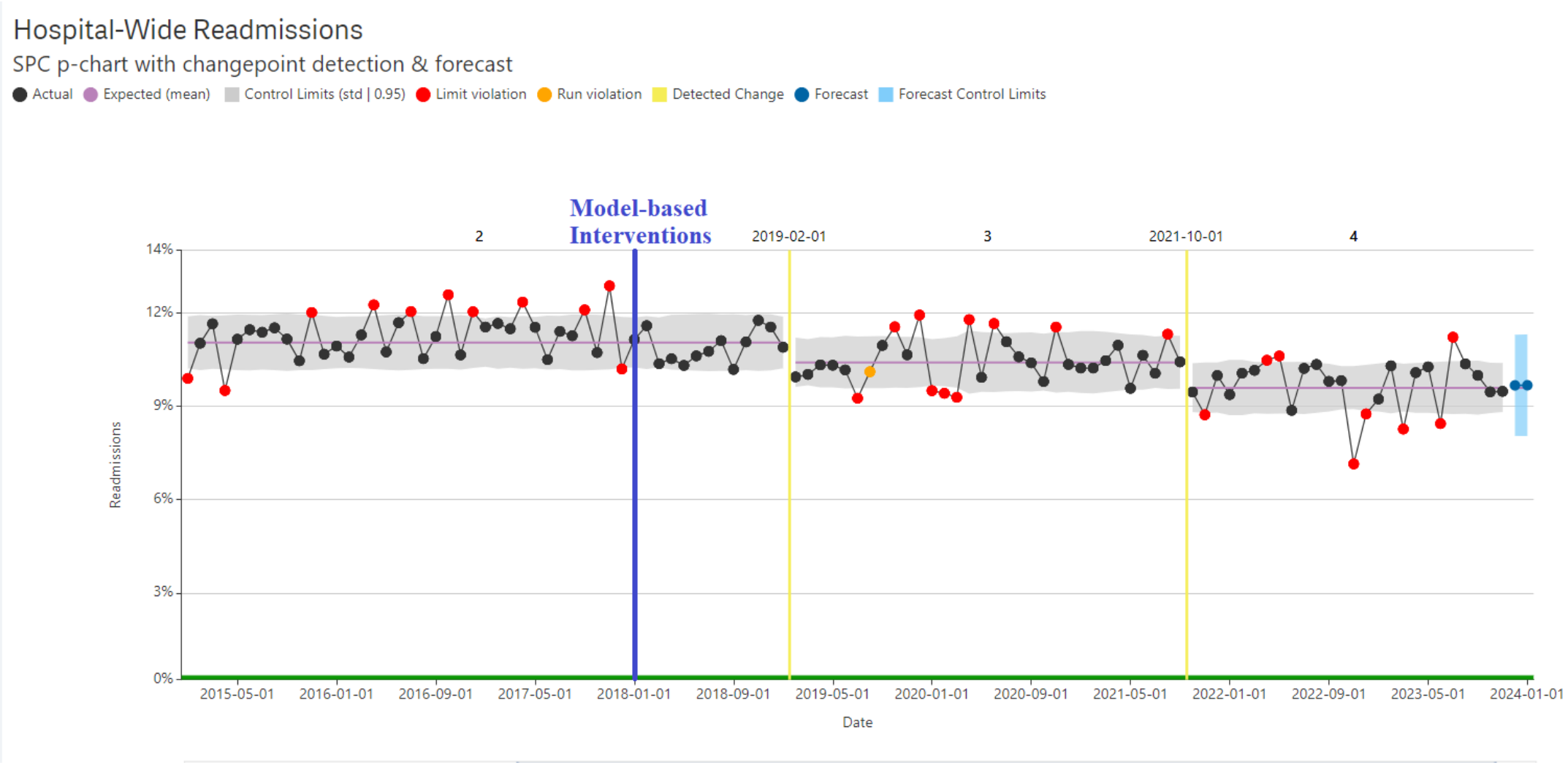
Interventions for all patients

- Medication delivery
- Social Determinants of Health
- Follow up appointments
- Discharge phone calls

Model-driven interventions

Time	Interventions	Low Risk	Moderate Risk	High Risk
DC Planning on Admit	Dedicated DC Navigator Nurse[i,ii,iii] *RN who focus on health literacy & education including assessments & teach-back before DC*		√	√
24 hours Post-DC Hand-off(s)	Patient Care Coordinator (PCC)[iv] or Insurance Care Manager	√		
	Outpatient Specialty Care Manager (CM)[v]		√	√
	Outpatient Specialty Social Worker (SW)[vi]		√	√
5-10 Days Post-DC	Telehealth appt with CM/SW[vii] *Verifies proper use of DME, medications & f/u appt including authorizations*		√	√
12-30 Days Post-DC Point of Contact	Patient Care Coordinator (PCC) or Care Manager Assistant (CMA)[viii] *Staff returns phone calls and follows up on outstanding issues*	+	++	√
	Care Manager/Social Worker[ix] *Licensed professional who returns phone calls and follows up on outstanding issues including alignment with community resources*			

All-cause Readmission Rates



Reduced Readmissions Saves $2.6M During First 24 months of Deployment

Readmission to the hospital within 30 days of discharge costs the US health system over $50B annually. Reducing readmissions is a key indicator of the quality of care and is critical to a health system's financial well-being.

Committed to reducing preventable hospital-wide readmissions, Children's Hospital of Orange County leveraged its data and analytics to significantly reduce its readmission rate, positively impacting patient lives and reducing costs.

$2.4M reduced costs, a result of **422 readmissions avoided** over 16 months.

11% relative reduction in hospital-wide readmission rate.

Accelerating Rare Disease Research Using Machine Learning and Artificial Intelligence

– A case study of juvenile dermatomyositis

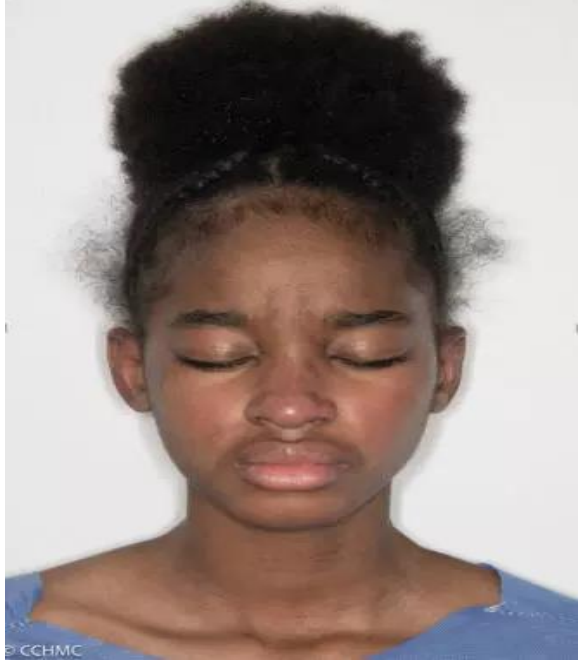
CHOC Research
GO BEYOND

Juvenile dermatomyositis

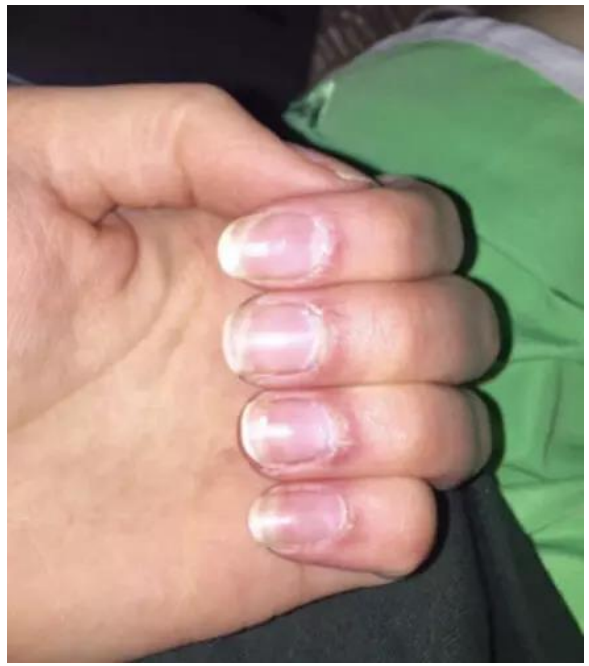
Juvenile dermatomyositis (JDM) is rare autoimmune diseases in which the body's immune system attacks its own cells and tissues resulting in both weak muscles and skin rash. Muscle weakness can be debilitating. The disease has no cure. Approximately two to four children in a million are diagnosed with JM (broader than JDM) each year in the United States.

Evidence of disease activity include:

- Muscle weakness
- Skin rash
- Fatigue
- Fever in some cases






[What is Juvenile Myositis? - Cure JM Foundation]

Treatment

There is no treatment, but medications can help alleviate symptoms. It involves administration of corticosteroids and methotrexates.

Disease activity scores have been developed

Expensive labs and biomarkers can be indicative of disease activity

Nailfold capillaroscopy is taken to assess damage to capillary bed of fingernails

The challenge

CHOC Research
GO BEYOND

The challenge

- Biomarkers for idiopathic inflammatory myopathies:
  - are difficult to identify
  - may involve expensive laboratory tests
- Nailfold capillaroscopy (NFC) are easily obtained
  - Can it be used to estimate disease activity?
  - Are there regions of the NFC that are informative of disease activity unknown to clinicians?
  - Can we predict resolution of symptoms?

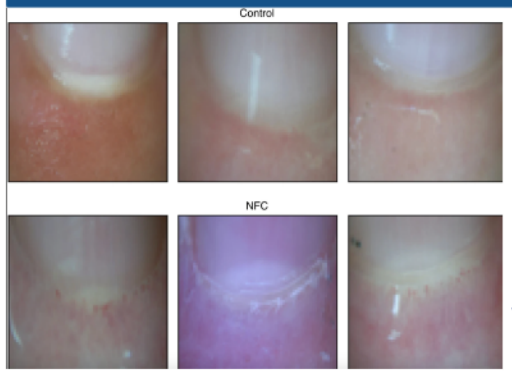
The approach

CHOC Research
GO BEYOND

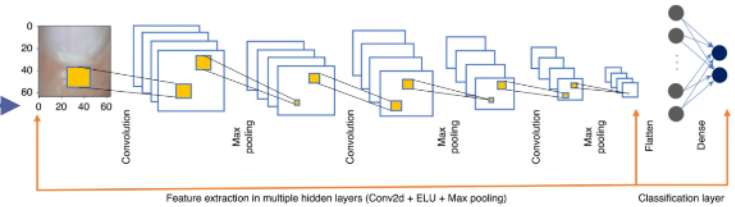
The approach

- We will assess the potential for artificial intelligence (AI) with deep learning to differentiate patients with Juvenile Dermatomyositis (JDM) from healthy controls using nailfold capillaroscopy (NFC) images
- Can active and inactive disease states be classified accurately among children with JDM?

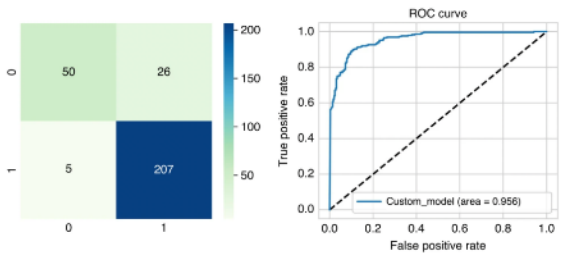
Artificial intelligence for nailfold capillaroscopy analyses – a proof of concept application in juvenile dermatomyositis
Hosseinzadeh Kassani, Ehwerhemuepha and Pachman et al. 2023



Separated into training/ testing sets using 5 fold cross validation



Images from Lurie's Children Juvenile Dermatomyositis (JDM) registry database
n=111 with JDM
n=31 controls



Model trained using a Deep Neural Network Architecture named Nailfold Capillaroscopy (NFC)-Net at Children's Hospital of Orange county

Ann & Robert H. Lurie Children's Hospital of Chicago®

CHOC

Results
NFC-net could accurately predict children in Dataset with JDM versus those who did not/ predicted disease severity

Limitations
Small overall dataset as result of rare disease and no external validation dataset

Funding gratefully acknowledged by authors and listed under article at DOI: https://doi.org/10.1038/s41390-023-02894-7

Pediatric RESEARCH

JDM: Explainability using integrated gradients

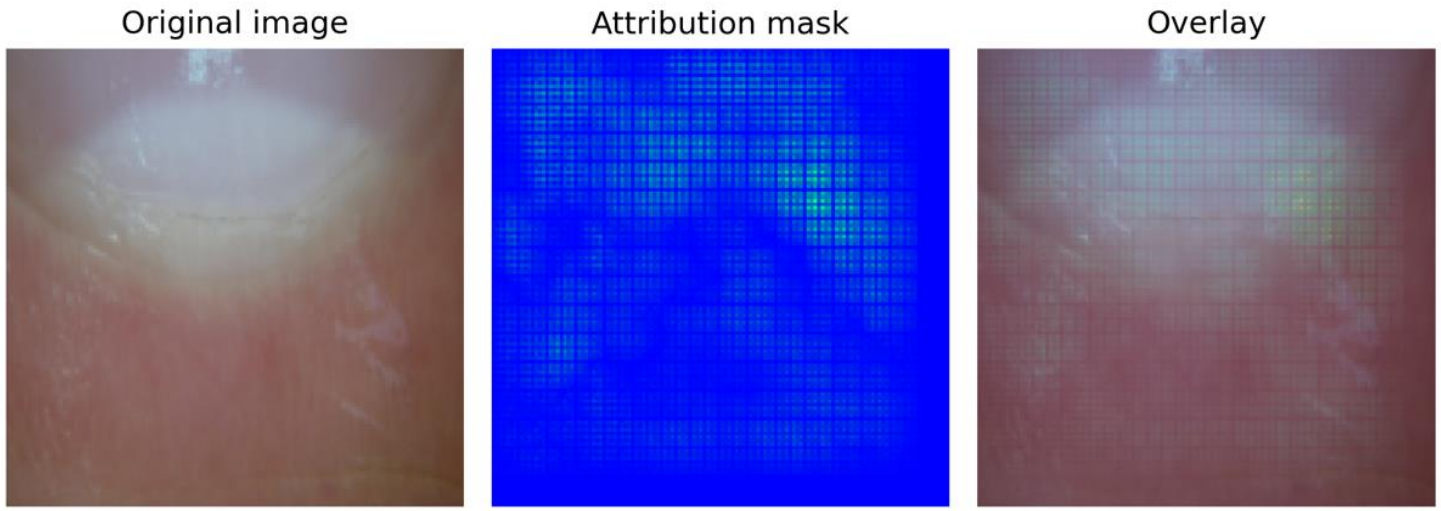


Fig. 4. Attribution map by integrated gradients. The baseline is created first; then an attribution map is calculated by integrated gradient and overlaid on the original image.

Other studies

CHOC Research
GO BEYOND

Submitted to the Annals of Applied Statistics

DYNAMIC TOPIC LANGUAGE MODEL ON HETEROGENEOUS CHILDREN'S MENTAL HEALTH CLINICAL NOTES

BY HANWEN YE[1,a], TATIANA MORENO[2,c], ADRIANNE ALPERN[2,d], LOUIS EHWERHEMUEPHA[2,e] AND ANNIE QU[1,b]

[1]*Department of Statistics, University of California, Irvine,* [a]*hanweny@uci.edu;* [b]*aqu2@uci.edu*

[2]*Children's Hospital of Orange County,* [c]*Tatiana.Moreno@choc.org;* [d]*AAlpern@choc.org;* [e]*LEhwerhemuepha@choc.org*

Mental health diseases affect children's lives and well-beings which have received increased attention since the COVID-19 pandemic. Analyzing psychiatric clinical notes with topic models is critical to evaluate children's mental status over time. However, few topic models are built for longitudinal settings, and they fail to keep consistent topics and capture temporal trajectories for each document. To address these challenges, we develop a longitudinal topic model with time-invariant topics and individualized temporal dependencies on the evolving document metadata. Our model preserves the semantic meaning of discovered topics over time and incorporates heterogeneity among documents. In particular, when documents can be categorized, we propose an unsupervised topics learning approach to maximize topic heterogeneity across different document groups. We also present an efficient variational optimization procedure adapted for the multistage longitudinal setting. In this case study, we apply our method to the psychiatric clinical notes from a large tertiary pediatric hospital in Southern California and achieve a 38% increase in the overall coherence of extracted topics. Our real data analysis reveals that children tend to express more negative emotions during state shutdowns and more positive when schools reopen. Furthermore, it suggests that sexual and gender minority (SGM) children display more pronounced reactions to major COVID-19 events and a greater sensitivity to vaccine-related news than non-SGM children. This study examines the progression of children's mental health during the pandemic and offers clinicians valuable insights to recognize the disparities in children's mental health related to their sexual and gender identities.

Computational and Data Sciences, Research Institute

Ricardo Aguilar, MS, RCS I, Applied and Causal Inference in Machine Learning

Comparative Effectiveness of Hydroxyurea and Voxelotor for Treatment of Sickle Cell Disease Using Target Trial Emulation

- Aim: To compare the effectiveness of hydroxyurea alone, voxelotor alone, and the concomitant use of both drugs

Voxelotor costs
$100,000 per year

https://www.pharmacytimes.com/view/daily-medication-pearl-voxelotor-oxbryta-

Hydroxyurea costs
$1,000 per year

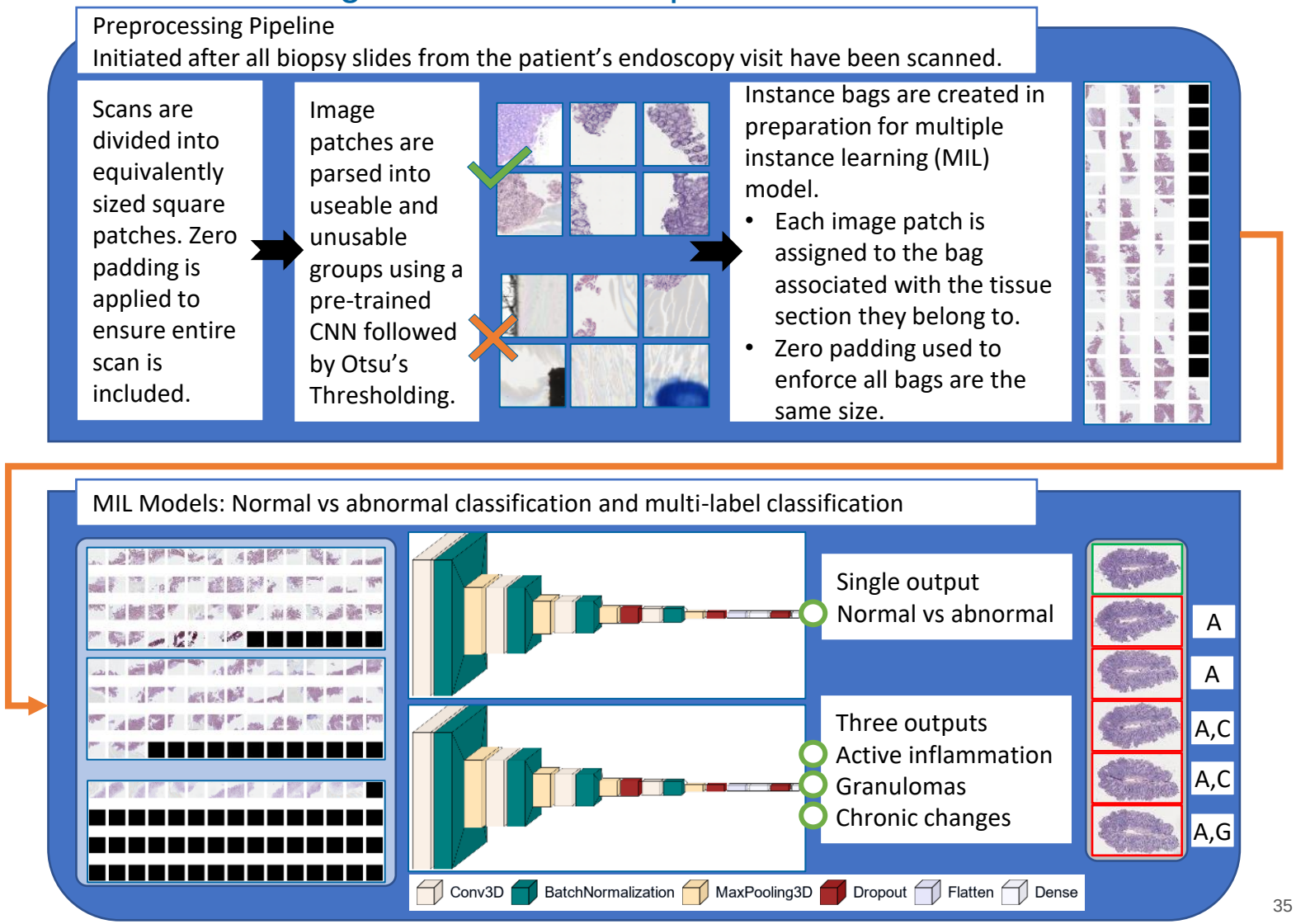
https://www.manovaistine.lt/private/uploads/images/products/hydrea-500mg-caps-n100.png

Computational and Data Sciences, Research Institute

Chloe Martin-King, PhD, RCS II, Computer vision and image restoration/interpolation

- Aims:
  - Classify each tissue section on a whole slide as normal or abnormal.
  - Subclassify abnormal sections as containing active inflammation, granulomas, chronic changes.
  - Assess steroid-free remission

- Clinical PI: Dr. Kenneth Grant, Gastroenterology



Lu *et al. BMC Medical Research Methodology* (2022) 22:181
https://doi.org/10.1186/s12874-022-01665-y

BMC Medical Research Methodology

RESEARCH

Open Access

A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance

Hongxia Lu[1], Louis Ehwerhemuepha[1,2] and Cyril Rakovski[1*]

PEDIATRICS®

Article Navigation

ARTICLES | SEPTEMBER 15 2022

Epidemiology of Neonatal COVID-19 in the United States

Open Access

Joan Devin, BSc; Rachel Marano, MD; Michel Mikhael, MD; William Feaster, MD; Terence Sanger, MD, PhD; Louis Ehwerhemuepha, PhD

Address correspondence to Louis Ehwerhemuepha, PhD, Computational Research, Children's Health of Orange County, 1201 West La Veta Ave, Orange, CA 92868. E-mail: lehwerhemuepha@choc.org

Pediatrics (2022) 150 (4): e2022056297.

https://doi.org/10.1542/peds.2022-056297 **Article history**

Computer vision and NLP projects

Summarization of PICU notes using large language models

Analyzing the risk of bias in large language models

Predicting pediatric-origin of headaches and migraines using NLP feature extraction for discovery of novel risk factors and predictors

Predicting CPAP failure, neonatal sepsis, and necrotizing enterocolitis using chest x-rays

Patent ductus arteriosus (PDA): Two studies using dual attention convolutional neural network and lightweight models with custom loss functions

Prediction and classification of velopharyngeal insufficiency: 3 projects including combination of CNN and RNN to capture temporal dependencies

Cortical dysplasia: multiple projects to predict cortical dysplasia using MRI images

AI is checking notes and raising standards in paediatric care

Forward-thinking hospital teams are harnessing the power of big data, machine learning and AI to drive research, improve patient outcomes and transform paediatric care.

Produced by

nature research custom media

CHOC

SAVE THE DATE!

PEDIATRIC & LIFESPAN DATA SCIENCE CONFERENCE

Thursday & Friday, April 10-11, 2025

Scan to learn more

Agenda - Pediatric and Lifespan Data Science Conference

Thursday, April 10

Time	Description
7:00 - 8:00 AM	Check-in
8:00 - 8:15 AM	Welcome Note
8:15 - 8:45 AM	Keynote Address
9:00 - 9:15 AM	Conference Outcomes Update
9:15 - 9:45 AM	Sponsored Coffee Break
9:45 - 10:45 AM	Generative Artificial Intelligence (AI) & Machine Learning Algorithms Towards Precision Medicine
11:00 - 11:45 AM	Role of Multimodal Electronic Health Records (EHR) Data & Health Equity in Precision Medicine
11:45 - 12:45 PM	Poster Presentations
12:30 - 1:30 PM	Sponsored Lunch

2:15 - 3:00 PM	Predicting Suicide Among Youths & Adolescents
3:15 - 4:00 PM	Proven Impact and Challenges of AI in Pediatric & Adult Medicine
4:15 - 4:45 PM	Abstract Podium Presentations
6:00 - 9:00 PM	Rooftop Cocktail and Dinner

Friday, April 11

Time	Description
7:00 - 8:00 AM	Breakfast
8:00 - 8:05 AM	Welcome Address
8:05 - 8:50 AM	Precision Medicine for Complex Medical Patients - Patient & Family Perspectives
9:00 - 9:45 AM	Hospital Administration, AI, & Advanced Analytics for Clinical Care Delivery
9:45 - 10:00 AM	Sponsored Coffee Break
10:00 - 10:45 AM	Health Insurance in the Era of AI & Advanced Analytics
11:00 - 11:25 AM	Driving Change Using Data & Improvement Science
11:25 - 11:45 AM	Datathon Updates & Presentations

CHOC Research
GO BEYOND.

High School AI Education Program

- Event to teach high school students data science and AI applications through fun activities
- Includes reinforcement learning to program a race car in collaboration with AWS
- Health data challenge
- Awards
- Prior to actual data science conference

LONG LIVE CHILDHOOD

Contact: Louis Ehwerhemuepha / lehwerhemuepha@choc.org

CHOC Research
GO BEYOND