# Life and Health Risk Assessment Using Non-Traditional Data using Deep Learning

LYDIA.ai

Anthony Lee, Co-Founder & CEO of Lydia Al 2024-10-10

## Anthony Lee Co-founder & CEO

LYDIA.ai

R&D in machine learning since 2015 from University of Toronto

### **Co-inventor**

- Method and system of using hierarchical vectorisation for representation of healthcare data (2021)
- Method and system for mapping text phrases to a taxonomy (2020)
- System and method for outputting groups of vectorized temporal records (2018)

Young Scholar

Innovation Award





2022 Credit Suisse Social Impact Award

Tatler Gen T. 2022 Power & Purpose Leaders of tomorrow

# At Lydia AI, we use machine learning to unlock new sources of data to make risk predictions



#### Our Mission

To protect the health & prosperity of the next billion people

Confidential

# Lydia AI risk scoring spans multiple dimensions of health, from clinical to physical activity

### **Clinical Score**

Factors in medical history and health predictions

### **Activity Score**

Factors in activity history based on wearable data

## Lifestyle Score

Factors in lifestyle habits (i.e., alcohol, smoking)

### **Mental Health**

Determines major depression based on wearable data



## Bridging the gap between AI/ML and traditional insurance



# A history of ML in insurtech: risk scoring using external data



## A history of ML in insurtech: RGA & Transunion Credit based Mortality



### Developed the US TrueRisk Life Model

- Predicts credit-based mortality for life insurance
- Core model built on 40 million lives and over 3 million deaths
- Scores validated on 18 million lives
- Study shows 5 times segmentation (96-100 compared to 1-5)



Source: 2018 SOA Predictive Analytics Seminar Presentation

# All risk score model building behaves in a similar framework







Model training

Retrain models on localized data source



### Validation dataset

Create validation dataset, representative of target population, to assess model

# How does what we learn from the RGA credit-mortality score help us?

Risk scoring offers more granular risk segmentation

Risk scoring often indicates value even above age of insurability

### Use cases from risk scoring include:

Underwriting triage



Cross-sell / upsell existing in-force

- Faster policy approvals
- Modifying UW rules

# Model Validation – Insured Lives Study

Small Face Whole Life Mortality Study

#### Details

- Includes Whole Life products < \$100k face; most of this business is
- under \$25k or \$50k
- Issue Ages < 70</li>
- Scores above 90 are further split out

#### **Results**

- Mortality about 6 times higher for worst scores
- Segmentation at higher scores for this business
- 14% of exposure & 29%
- of claims have score > 95 • > 10% of the claims have
- a score of 100Value also seen beyond
- age 70



350%

300%

250%

200%

100%

50%

Overall Mortality Issue Age < 70 Second Sec Multiple risk scores have been implemented by (re)insurers to focus specifically on pre-approvals, preferred kick-out, and UW relaxation in the US

	<b>WexisNexis</b> Risk Solutions Risk Classifier (LNRC)	A Quest Diagnostics Company Clinical Lab Data	ExamOne* A Quest Diagnostics Company LabPiQture	Milliman IntelliScript® Rx Score (Irix)	
Data Type	Credit score motor vehicle records public records	Historical lab test results	Three sub-scores: Prescription history, Lab test results, Diagnosis codes	Prescription data	
Data Size	2.2M de-identified gen pop & insured 5 years, 51k deaths	83k insured lives, 7 years from FUW,	4.2M de-identified health data, 63k deaths	25m gen insured lives, 468k deaths, 104m exposure years	
Validation	Munich RE E Gen Re. PartnerRe	Munich RE <b>RGA</b>	Munich RE hannover <b>re</b>	Munich RE	

## Milliman Rx Score (Irix) 2.0: Drug Data Only

Each life entered the study between the first quarter of 2005 and the last quarter of 2016. **Deaths were sourced from the Social Security Death Master File** and third party proprietary databases.

The study population is comprised of 468,491 deaths out of 104 million exposed life-years.

The expected mortality basis was taken from the **2015 VBT primary select** and ultimate ANB tables split by age and gender with a 1 percent mortality improvement



X-axis is Milliman Rx Score; lines are age buckets; Y-axis is relative A/E

All age groups follow the same pattern, where mortality risk increases as Rx scores increase



X-axis is age group; y-axis is relative A/E. % of people with corresponding Milliman Rx Score 2.0

We see that the distribution of scores is about the same across age groups.

## **ExamOne LabPiQture: Clinical Lab Data with Medical Data**

Munich Re performed a mortality study on a final, clean version of the ExamOne data set with 62,415 deaths. The expected basis was the Society of Actuaries (SOA) 2015 Select and Ultimate Valuation Basic Table (VBT) with mortality improvement from 2015 onwards and dis-improvement prior to 2015.



X-axis risk scoring bands, with lower bands = lower risk Y-axis is risk of mortality relative to cohort; lines represent mortality trends between scores

Among the component scores, clinical lab scores are best at segmenting mortality followed by prescription history and diagnosis scores.



X-axis risk scoring bands, with lower bands = lower risks Y-axis is risk of mortality relative to cohort.

All age groups follow the same pattern, where mortality risk increases as final Score increases.

## MassMutual LifeScore360: Commercial UW + Claims Input

MassMutual has a consolidated, digital record of nearly one million applications for which a lab test was ordered during 1999–2014. After removing applications with a high degree of missing values, typically incom- plete or withdrawn applications, this reduces to 908k records with 9.16M exposure years and 15.7k observed deaths.



X-axis risk scoring bands, with higher numbers = higher risk Y-axis is proportion of population with corresponding age group

Among the component scores, clinical lab scores are best at segmenting mortality followed by prescription history and diagnosis scores.



X-axis risk scoring bands, with % of population with BMI from 20-to-45: Y-axis is proportion of people in each corresponding Life Score decile

Top figure describes of heart condition as a function of life score. The gradually decreasing in between. Distribution of BMI as a function of life score. The highest scores have a greater proportion of healthy-range BMI. As the score decreases, the proportion of upper and lower BMI extremes gradually increases. 12

## Methodological differences have moved from feature engineering to few shot or zero shot learning

New tools offer advantages when data is complex and labels are few.

#### **Traditional Machine Learning**

Humans selected limited features to use for predictions



### Pre-trained Language Models

Models automatically leverage extensive features for predictions





### Challenges with existing tools

Dependency on large labelled datasets

Data imbalance and distribution issues

Complexity of feature engineering



Complex data integration capabilities

Automated feature extraction

Zero-shot and few-shot learning

## AI/ML Methods have evolved significantly since the 90s



# Deep learning methods versus traditional methods to learn multiple levels of features



**Traditional Machine Learning** 

## For example Lydia AI's activity score looks at activity data from Apple Health

Correlation of health score to



Correlation of health score to age





<sup>1</sup>Based on gender-age specific cut-points of maximal oxygen consumption (VO2 max) measurements

©2023 Lydia AI. Knowtions Research. Proprietary and to be used only with permission

Risk Bir

We maintain the tradition of grouping people by "deciles" and looking at correlations between "deciles" and clinical indices

• Box Plot and Interquartile Range (IQR)



• Sample Correlated Feature maybe "Diagnosis Code Count"

Total diagnosis counts (non-unique) across patient history, inclusive of primary diagnosis and secondary diagnosis codes

# We can understand each decile by its respective incidence rates as benchmarked with industry or country incidences (i.e., mortality, hospitalization)



Age	Male Hospitalization rate (qx)									
	Top 10%	Top 20%	Top 30%	Top 40%	Top 50%	Top 60%	Top 70%	Top 80%	Top 90%	Top 100%
0-4	0.211	0.301	0.342	0.364	0.384	0.378	0.368	0.358	0.357	0.368
5-9	0.002	0.016	0.037	0.056	0.079	0.086	0.096	0.104	0.118	0.143
10-14	0.012	0.026	0.037	0.043	0.050	0.052	0.054	0.057	0.062	0.071
15-19	0.028	0.049	0.066	0.070	0.073	0.075	0.076	0.078	0.082	0.086
20-24	0.029	0.053	0.070	0.082	0.084	0.087	0.088	0.092	0.095	0.098
25-29	0,053	0.076	0.081	0.084	0.068	0.092	0.095	0.100	0,106	0.112
30-34	0.045	0.082	0.092	0.097	0.104	0.112	0.119	0.128	0.138	D.152
35-39	0.050	0.074	0.080.0	0.088	0.092	0.100	0.108	0.118	0.129	0.146
40-44	0,036	0.064	0.075	0.084	D.089	0.094	0.102	0.111	0.123	0.142
45-49	0.040	0.068	0.077	0.087	0.095	0.102	0.109	0.118	0.129	0.147
50-54	0.041	0.059	0.074	0.087	0.096	0.104	0.113	0.126	0.140	0.157
55-59	0.039	0.063	0.079	0.093	D.109	0.123	0.137	0.150	0.167	D.188
60-64	0.036	0.058	0.077	0.088	0.105	0.121	0.140	0.163	0.192	0.219
65-69	0.026	0.050	0.067	0.086	0.106	0.123	0.146	0.179	0.211	0.245
70-74	0.009	0.044	0.070	0.090	0.116	0.147	0.172	0.208	0.243	0.284
75-79	0.029	0.047	0.070	0.092	0.122	0.159	0.199	0.240	0.283	0.334
80+	0.062	0.084	0.114	0.163	0.204	0.238	0.281	0.331	0.384	0.443
Overall	0.039	0.065	0.078	0.088	D.098	0.107	0.118	0.132	0.150	0.173



Age	Male Hospitalization rate (qx)									
	0.00-0.10	0.10-0.20	0.20-0.30	0.30-0.40	0.40-0.50	0.50-0.60	0.60-0.70	0.70-0.80	0.80-0.90	0.90-1.00
0-4	0.211	0.431	0.445	0.417	0.414	0.361	0.338	0.331	0.351	0.408
5-9	0.002	0.059	0.092	0.097	0.132	0.115	0.139	0.160	0.286	0.420
10-14	0.012	0.041	0.067	0.057	0.077	0.073	0.075	0,126	0.211	0.380
15-19	0.028	0.067	0.087	0.079	0.086	0.098	0.096	0.140	0.308	0.345
20-24	0.029	0.073	0.095	0.117	0.091	0,103	0.116	0.176	0.229	0.408
25-29	0.053	0.094	0:090	0.091	0.113	0.118	0.139	0.210	0.303	0.471
30-34	0.045	0.111	0.114	0.112	0.140	0.152	0.185	0.267	0.354	0.471
35-39	0.050	0.096	0.093	0.111	0.113	0.152	0.181	0.250	0.334	0.479
40-44	0.036	0.095	0.100	0.112	0.116	0.124	0.159	0.203	0.261	0.465
45-49	0.040	0.100	0.102	0.124	0.133	0.141	0.157	0.189	0.256	0.429
50-54	0.041	0.084	0.118	0.130	0.135	0,147	0,166	0.212	0.274	0.424
55-59	0.039	0.101	0.130	0.147	0.184	0.194	0.208	0.222	0.288	0.423
60-64	0.036	0.093	0.147	0.136	0.184	0.193	0.224	0.268	0.349	0.437
65-69	0.026	0.095	0.144	0.169	0.204	0.213	0.254	0.309	0.347	0.463
70-74	0.009	0.110	0.170	0.183	0.222	0.258	0.265	0.336	0.372	0.492
75-79	0.028	0.084	0.240	0.208	0.244	0.310	0.364	0.377	0.417	0.539
80+	0.062	0.120	0.223	0.393	0.359	0.393	0.415	0.476	0.541	0.624
Overall	0.039	0.095	0.109	0.120	0.143	0.160	0.193	0.249	0.323	0.456

# We see further complexity in ML models that deal with edge cases and large evolving data patterns

### Facial Recognition Example

Lee H, Grosse R, Ranganath R, Ng AY. Unsupervised learning of hierarchical representations with convolutional deep belief networks.





**Raw image details –** edges, blobs, contrasting lines, ...



Face parts – noses, eyes, eyebrows, ears,

cheeks....

# 





**Local features** – trends, spikes, waves, etc.

Lydia AI Uses 1-D Convolutional Neural Network

Mid-range patterns – combinations of local features that are useful for the predictive task at hand





Face models – translation invariant face archetypes



Long range patterns – signal patterns that occur on longer time scales and whose presence or absence allows the model to come to a decision

# Local features are found in activity signals when looking at steps, as opposed to being pre-engineered as rules by bioinformaticians



## Over 30,000+ dimensions of features

## Sample #66123

Age: 29 | Gender: Female

![](_page_20_Figure_2.jpeg)

Intensity
Avg Doily Intensity / Min

## Sample #67032

Age: 32 | Gender: Female

![](_page_21_Figure_2.jpeg)

## Sample #63753

Age: 33 | Gender: Male

![](_page_22_Figure_2.jpeg)

Intensity Avg Doity Intensity / Min

©2023 Lydia AI. Knowtions Research. Proprietary and to be used only with permission

# We now understand average number of steps, intensity of activity and consistency in activity pattern are all key factors in affecting the physical activity score

![](_page_23_Figure_1.jpeg)

On average, the higher average steps over a 7-day period will likely result in a better score

#### Case study Score bin: 0.90 - 1.00

![](_page_23_Figure_4.jpeg)

![](_page_23_Figure_5.jpeg)

Higher average intensity of activity (steps per min), will likely result in a better score

![](_page_23_Figure_7.jpeg)

Consistent patterns that show trends of high vs. low activity will likely result in a better score

![](_page_23_Figure_9.jpeg)

- Lower number of steps per week
- Lower average intensity of activity
- Lower range of movement pattern

![](_page_23_Figure_13.jpeg)

![](_page_23_Figure_14.jpeg)

- Lower number of steps per week
- Lower average intensity of activity
- Lower average intensity of activity
- Lower range of movement pattern

©2023 Lydia AI. Knowtions Research. Proprietary and to be used only with permission

We can even zero in on specific user segments to focus on life-stage specific needs

Japanese women from teens to 20s: Correlation between activity and PMS

Journal: BMC Sports Sci Med Rehabil. 2022; 14: 175.

### **Result Highlights:**

• A survey of 381 Japanese female university students (average age 20.44) found that those who engaged in more than 3,000 MET minutes of physical activity per week had lower overall PMS symptom scores, including lower physical and psychological symptoms, compared to those with less activity. \*3,000 MET minutes per week is equivalent to about 60 minutes of walking and 30 minutes of jogging per day, indicating moderate exercise levels.

 $Walking MET - \frac{minutes}{week} = 3.3 * walking minutes * walking days$  $Moderate MET - \frac{minutes}{week} = 4.0 * moderate - intensity activity minutes$ \* moderate - intensity days

 $Vigorous MET - \frac{min}{week} = 8.0 * vigorous - intensity activity minutes$ \* vigorous - intensity days

Total physical activity MET – minutes/week = sum of walking + moderate + vigorous MET – minutes/weekscores We can even zero in on specific user segments to focus on life-stage specific needs

Women in their 20s to 30s: Correlation between Activity and Sleep

Journal: Front. Public Health, 03 November 2022

### **Result Highlights:**

### **Physical Activity**

- **Moderate exercise** has a positive effect on reproductive health, while excessive exercise may negatively impact the menstrual cycle.
- Lack of exercise can lead to obesity, increasing the risk of polycystic ovary syndrome (PCOS).
- **Excessive exercise** can disrupt the hypothalamus-pituitary-gonadal axis (HPT axis), affecting hormone balance and sperm quality.
- **Sleep deprivation** can raise thyroid-stimulating hormone (TSH) levels, potentially causing irregular or absent menstruation.
- Short sleep duration in men is linked to reduced sperm quality and an increase in anti-sperm antibodies, which can damage healthy sperm.

### Sleep

Sleep deprivation can increase thyroid-stimulating hormone (TSH) levels, potentially causing irregular menstruation or amenorrhea.

In men, **shorter sleep duration is associated with reduced semen quality.** Studies have found that men who sleep less tend to have poorer sperm quality, and an increase in anti-sperm antibodies. These antibodies can damage healthy sperm, potentially leading to a decline in sperm quality. We can even zero in on specific user segments to focus on life-stage specific needs Women in their 30s to 40s: Correlation between activity and peri-pregnancy or postpartum depression

Journal: Journal of Affective Disorders 246 (2019)

**Result Highlights:** 

- Average participant age: 33.3 years (±2.9 years).
- **Effects of Exercise During Pregnancy**: Physical activity during pregnancy has been suggested to prevent postpartum depression and is also beneficial for managing depression symptoms during pregnancy.
- Effects of Exercise Postpartum: Physical activity after childbirth has been shown to help in both preventing and treating postpartum depression.
- **Recommended Exercise:** Moderate-intensity exercise for 30 minutes or more, 1 to 4 times a week, is recommended to reduce the risk of postpartum depression.

![](_page_26_Figure_7.jpeg)

Figure 2 CES-D score at the three time points. CESD-D, Center for Epidemiological Studies-Depression Scale; CG, control group; IG, intervention group.

### We can even zero in on specific user segments to focus on life-stage specific needs

Women in their 40s and 50s: menopause and postmenopause

#### Journal: International Journal of Women's Health

### **Result Highlights:**

# Participants were divided into three groups for comparison:

Group A: Less than 30 minutes of physical activity per day.

Group B: 30 to 60 minutes of physical activity per day.

roup C: More than 60 minutes of physical activity per day.

Results showed that more than 60 minutes of moderate-intensity physical activity per day led to significant effects, including: Reduction in menopausal symptoms. Improved quality of life, particularly in psychological and social aspects. Decreased menopausal symptoms correlated with potential weight loss benefits.

![](_page_27_Figure_9.jpeg)

# The fun part of operating a health AI is forcing everyone to exercise to see real-world results

Scoring is based on activity consistency and changes:

- Maintaining the same activity level results in a 5-point increase.
- Moving up to a higher level of activity results in a 10-point increase per level.
- Decreasing the activity level results in a 10-point deduction.

#### The default score is set at 40, with a minimum of 0 and maximum of 99.

Does not meet quals, score:40 Moderate: 0 min, Vigorous: 0 min Week 2: Does not meet goals, score:40 Moderate: 0 min, Vigorous: 6 min Meinel Re-Does not meet quals, score:40 Maderate: 0 min, Vigorous: 69 min Does not neet goals, score:40 Moderate: 83 min, Vigorous: 8 min not meet goals, score:40 Moderate: 0 min, Vigorous: 33 min ets viporous goal, score:70 Moderate: 5 min, Viporous: 93 min Neets viporous goal, score:75 Moderate: 6 min, Vigorous: 141 min Week 8: Meets vigorous goal, score:80 Moderate: 0 min, Vigorous: 530 min Lvdian #1 300 250 200 150 100 20000 40000 60000 80000 minutes

![](_page_28_Figure_7.jpeg)

Neek 1: Neets both noderate and vigorous goals, score:80 Moderate: 200 min, Vigorous: 210 min Neek 2: Neets both noderate and vigorous goals, score:80 Moderate: 240 min, Vigorous: 210 min Neek 3: Neets both noderate and vigorous goals, score:80 Moderate: 240 min, Vigorous: 127 min Neek 4: Meets both noderate and vigorous goals, score:80 Moderate: 300 min, Vigorous: 137 min Neek 5: Neets both noderate and vigorous goals, score:80 Moderate: 350 min, Vigorous: 267 min Neek 5: Neets both noderate and vigorous goals, score:80 Moderate: 350 min, Vigorous: 267 min Neek 5: Neets both noderate and vigorous goals, score:80 Moderate: 350 min, Vigorous: 200 min Neek 7: Neets both noderate and vigorous goals, score:80 Moderate: 240 min, Vigorous: 200 min Neek 7: Neets both noderate and vigorous goals, score:80 Moderate: 440 min, Vigorous: 200 min Neek 7: Neets both noderate and vigorous goals, score:80 Moderate: 440 min, Vigorous: 200 min

![](_page_28_Figure_9.jpeg)

## **Risk scoring and activity goals and collected data show strong correlation over time** (*p*<0.001)

#### Scoring is based on activity consistency and changes:

- Maintaining the same activity level results in a 5-point increase.
- Moving up to a higher level of activity results in a 10-point increase per level.
- Decreasing the activity level results in a 10-point deduction.

### The default score is set at 40, with a minimum of 0 and maximum of 99.

![](_page_29_Figure_6.jpeg)

# Understanding how to make customers healthier through data will be one of the tasks for health actuaries in the next three years

![](_page_30_Figure_1.jpeg)

## Summary

### • The Future of Risk Assessment:

- A shift towards personalization and real-time analytics driven by advanced ML/AI techniques.
- Integration of non-traditional data opens new avenues for understanding and mitigating risks.

### • Ethical and Collaborative Approach:

- Emphasis on data privacy, security, and ethical use of AI.
- Building partnerships within the industry to enhance collective capabilities.

## • Our Call to Action:

- Encourage stakeholders to embrace innovative technologies and collaborative efforts.
- Explore opportunities to integrate causal inference methodologies for deeper insights.

## Setting the stage for causal inference in risk assessment

### • The Importance of Understanding Causality:

- Recognizing not just correlations but causal relationships enhances risk assessment accuracy.
- Facilitates the design of effective interventions and personalized strategies.

### • Complementing Our Approach:

- While Lydia AI focuses on predictive analytics using non-traditional data, understanding causality adds depth to our insights.
- Collaboration with experts in causal inference can lead to more robust risk models.

### • Introduction to Upcoming Topics:

- The next presentation by Kaz from Sumitomo Life Insurance will delve into *causal inference methods*.
- Discussions will cover foundational concepts, removal of biases, and practical applications in life insurance.

### • Working Collaboratively:

- Shared goals in leveraging data and technology for better risk assessment
- Developing a new generation of wellness programs and personalized insurance solutions **not just** about giving discounts but about rewarding current health discipline, incentivizing healthy and responsible behavior, and helping high-risk indviduals plan better