IOP Joint APP and HEPP Annual Conference 2025

Validation and Calibration of Machine Learning Models: From Particle Identification to Eye-Disease Prognosis

Robert McNulty

r.j.mcnulty@liverpool.ac.uk

Supervisors:

Dr. Nikolaos Rompotis, Prof. Monica D'Onofrio,

Prof. Yalin Zheng, Dr. Philip Burgess

UNIVERSITY OF LIVERPOOL   ATLAS EXPERIMENT   LIV.INNO   ARO Technology. Only better.

Introduction

- 3rd Year PhD Student at University of Liverpool (UoL)
- Apart of LIV.INNO CDT – joint funded with local industry partner (ARO)
  - PhD aimed at R&D projects connected to both ATLAS and ARO
- ATLAS: validation of GNN model for tau-lepton identification (co-developed with Dr. J Carmignani)
- CRiA (through ARO): validation of a time-distributed CNN model for eye-disease prognosis (developed by Dr. J Bridge)
- Today, I will be giving a brief overview of this work, highlighting connections and what has been learned from one field applied to the other

Brief Overview of τ-Leptons at ATLAS and Age-related Macular Degeneration

- τ_{had} BR > τ_{lep} BR (65% and 35%)
- τ_{had} decays = 1- or 3-prong (1 or 3 $\pi^{\pm}$'s, and maybe some π^{0}'s)

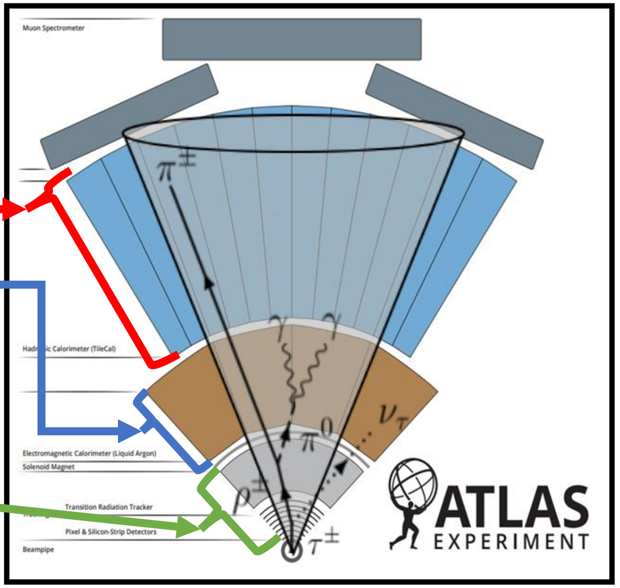
Hadronic Calorimeter
Obtains $\pi^{\pm}$ information

Electromagnetic Calorimeter
Obtains π^{0} information
(via $\pi^{0} \rightarrow \gamma\gamma$ and $\gamma \rightarrow e^{-}e^{+}$)

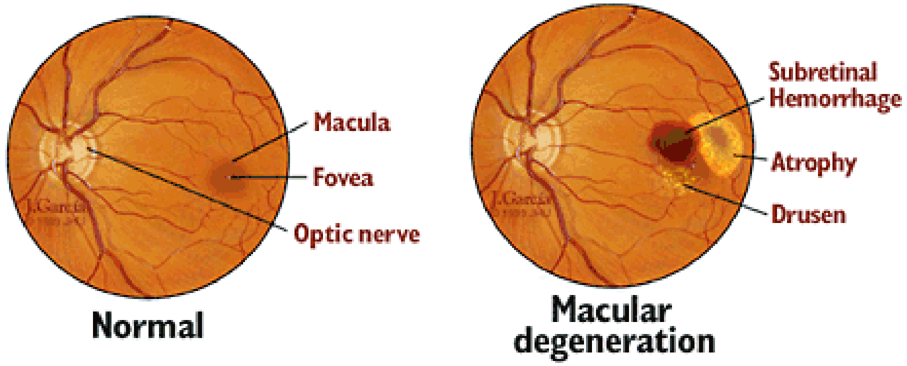
Tracking Detector
Collects charged particle track information, e.g., direction and position of $\pi^{\pm}$'s from τ-decay

τ_{had} Decay:

- Highly collimated – narrow cone
- Small cross-section
- Low multiplicity



- **Macula** = Small area at the centre of the retina responsible for central vision, most of colour vision, and finer details of sight
- Damage is **permanent**
- **Age-related Macular Degeneration (AMD)** = Common eye disease that blurs central vision
  - Occurs when aging causes damage to the macula
  - Leading cause of vision loss for older adults (typically 50+ years) – but doesn't cause complete blindness



Age-Related Eye Disease Study Dataset

- Data is from the (publicly available) Age-Related Eye Disease Study (AREDS) dataset [NCBI]
- Each eye has up to four retinal images taken at a clinical visit
  - Images have time stamp for respective visit and a marker for progression observed
- Longitudinal data – addition of time as a variable causes signal and background labels to change over time
  - E.g., in (b) the patient's eye progresses to nAMD at 5 years – this is the point where this event would be labelled with a "1" (signal)

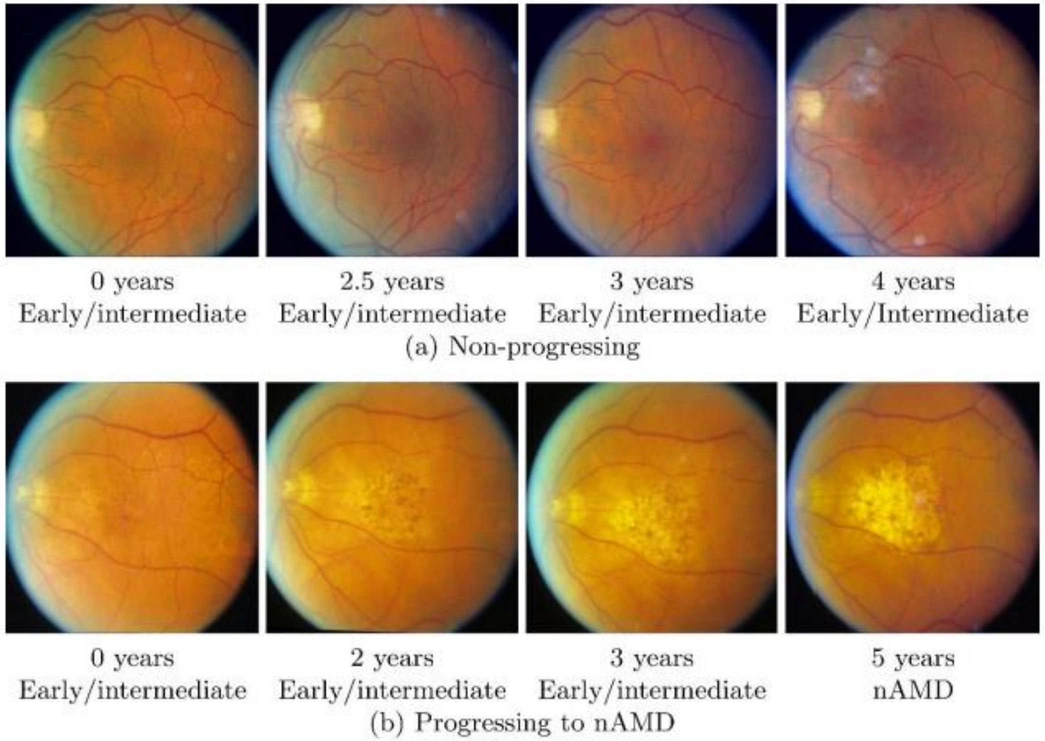


Figure available from [J. Bridge, 2022, PhD Thesis]

Problem and Approach

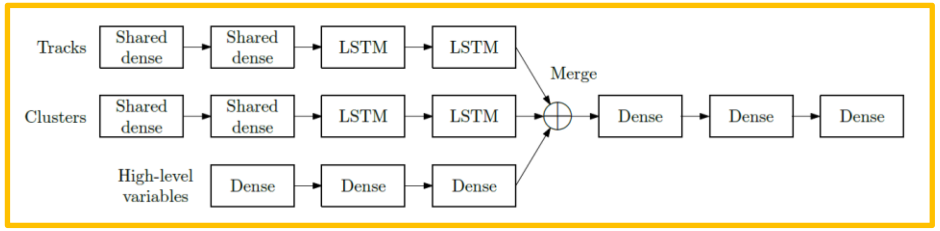
	τ_{had} Identification (and Decay Mode Classification)	AMD Progression

Problem	• τ_{had} signatures are "drowned-out" by QCD jets
	• Currently done via RNN (at ATLAS) [ATL-PHYS-PUB-2022-044]
	    • RNN inputs can be used as an image input to a GNN to perform tau-lepton identification
		• AMD progresses at different times for different people/eyes
		• Current approach is a time-distributed CNN (Survival Model) [J. Bridge, 2022, PhD Thesis]
		    • Uses real images as inputs
		    • Images are of the same eye (for each "event") over time
		• Classification problem, but now depends on time as an input/parameter

Approach	• To further study tau-lepton identification with a GNN based approach (TauJetGraphs):
	    • Requirement: to handle τ_{had} candidates with 1 and 3 tracks
		• To modify, extend, and finalise a Time-Distributed CNN (Survival Model)
		    • Requirement: to return a probability of a patient progressing by a given time, t

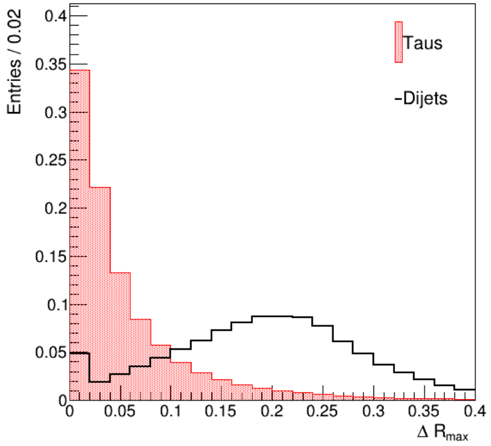
Tau ID – From RNN to GNN

Current: RNN [ATL-PHYS-PUB-2022-044]



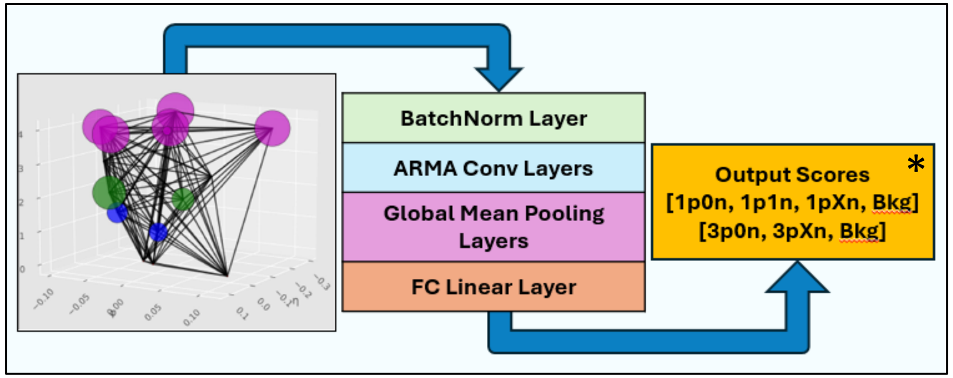
- RNN inputs: tracking, cluster, and high-level (global) jet variables
  - Can vary in length, but must be ordered in some way (e.g., by transverse momentum, p_T)
- GNN inputs: combination of input variables from ID-RNN and additional π^0 variables
  - Can be unordered sets with varying lengths

Example: 1-prong, ΔR



N.B.: 1- and 3-prong trained separately

Proposed: TauJetGraphs, GNN



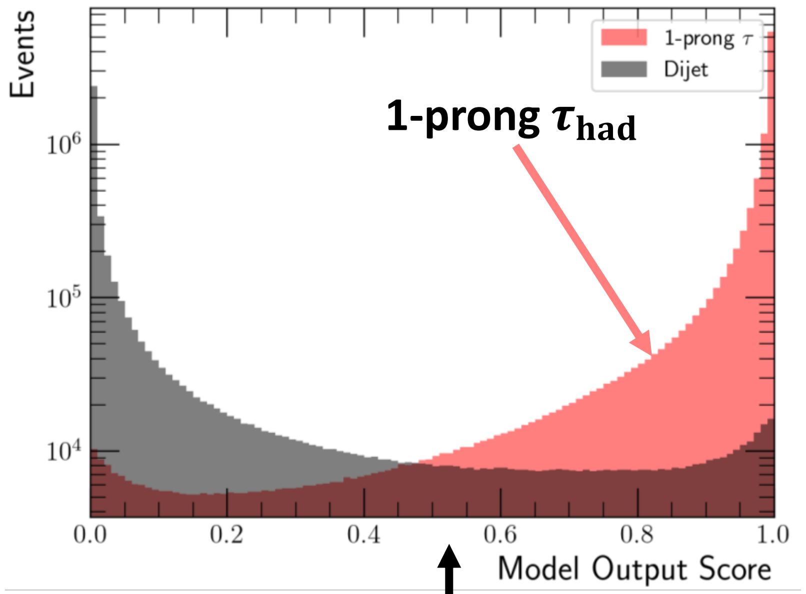
- Nodes = physics object, Layers = node label
- Nodes within a predefined distance ($\Delta \mathrm{R} \leq 0.4$) are connected by an edge, where

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

* Scores given for decay modes are summed for ID, e.g., $\mathrm{P(1p0n) + P(1p1n) + P(1pXn) = P(1-prong)}$

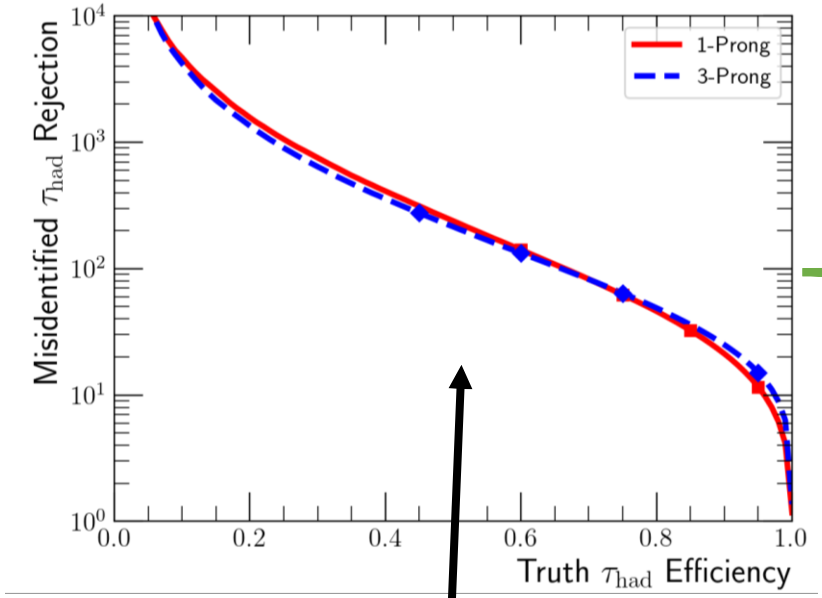
TauJetGraphs – Example of Outcomes* (ID)

*Based on Monte Carlo Simulations



Good signal-background separation

ROC Curve



3-prong performance shows similar performance to 1-prong

	1-prong			
Efficiency	60%	75%	85%	95%
Rejection	140	62	32	12

	3-prong			
Efficiency	45%	60%	75%	95%
Rejection	280	130	64	15

- $\text{Efficiency} = \frac{\text{True Positives}}{\text{True Positives+False Negatives}} (\times 100\%)$
- $\text{Background Rejection} = \frac{\text{False Positives+True Negatives}}{\text{False Positives}}$

Survival Models – Overview

- Survival Models: Predict survival of event for specified time interval
- Three main stages:
  1. **Feature Extraction (via a CNN)** – results in a feature vector for each image, F_N
  2. **Mixed-Effects (ME)** – accounts for missing images and times, x_i
     a) Clinical data can then be appended to the single vector
  3. **Time-distributed CNN (Survival Model)** – estimate the Survival Function, $S(t)$

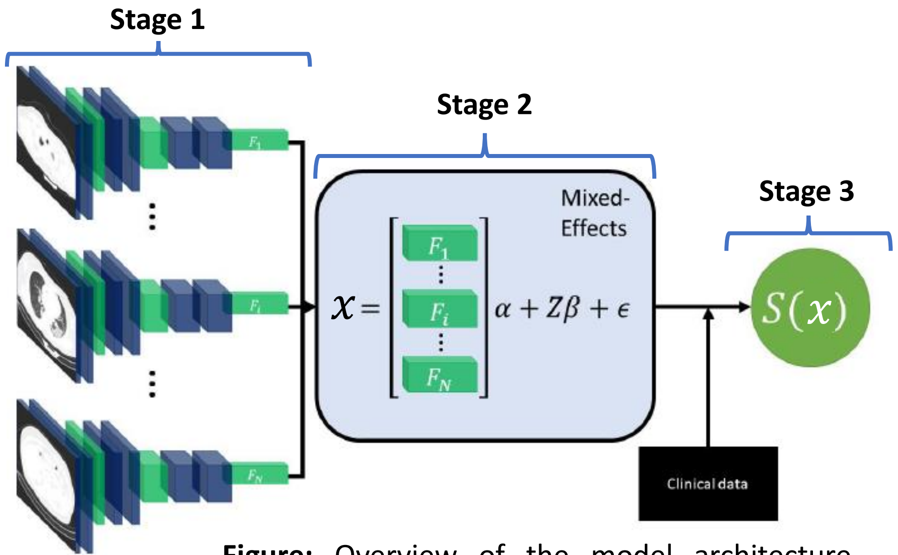
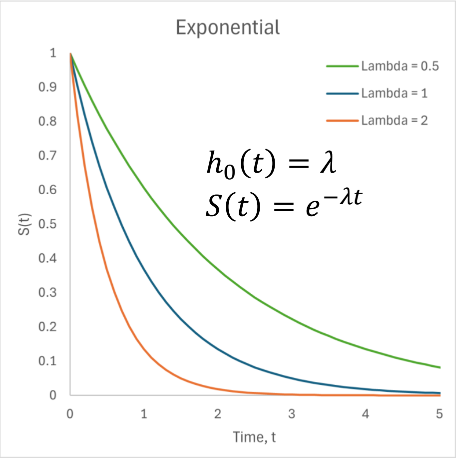
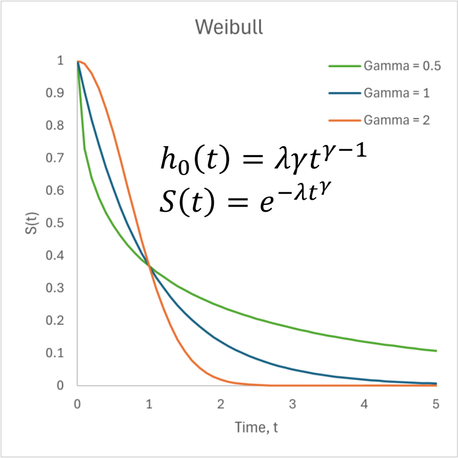
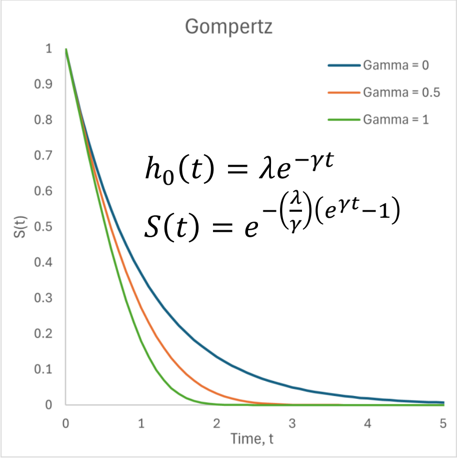


Figure: Overview of the model architecture. Image credit: Dr. Joshua Bridge. Figure available from [J. Bridge, 2022, PhD Thesis]

- Baseline Hazard Function: $h_0(t)$
- Hazard Function: $h(t) = h_0 e^{x\beta}$
- Survival Function: $S(t) = -\exp\left\{-\int_0^t h(u)\,dt\right\}$
- Survival Probability: $S(t) = P(T \geq t)$
- Failure Probability: $F(t) = P(T < t) = 1 - S(t)$







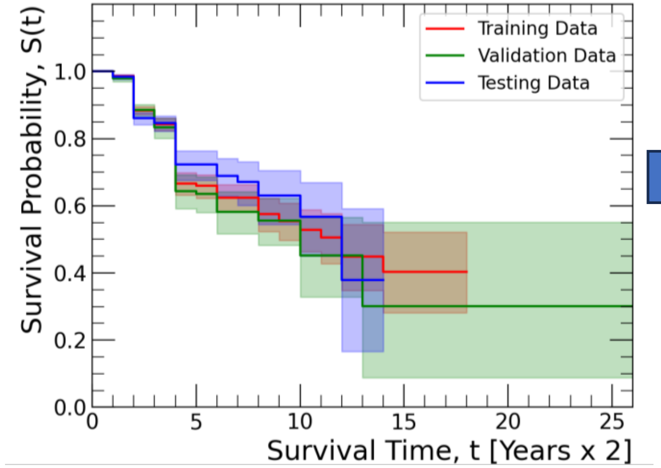
Survival Models – Kaplan-Meier Curves

Example – Exponential Baseline Hazard

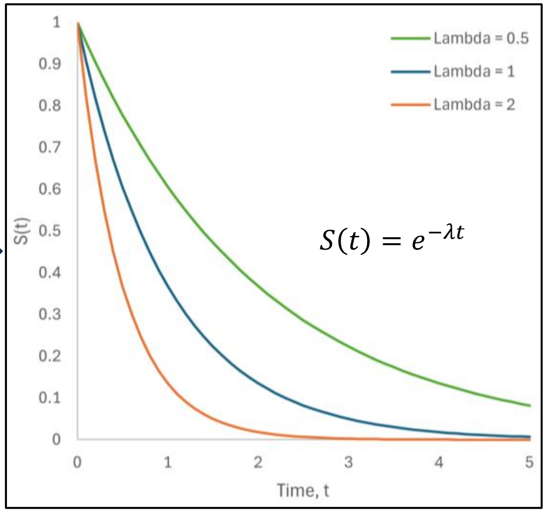
- Kaplan-Meier (KM) curves [doi: 10.4103/0974-7788.76794] are used to compare survival probabilities, which is given by (for KM):

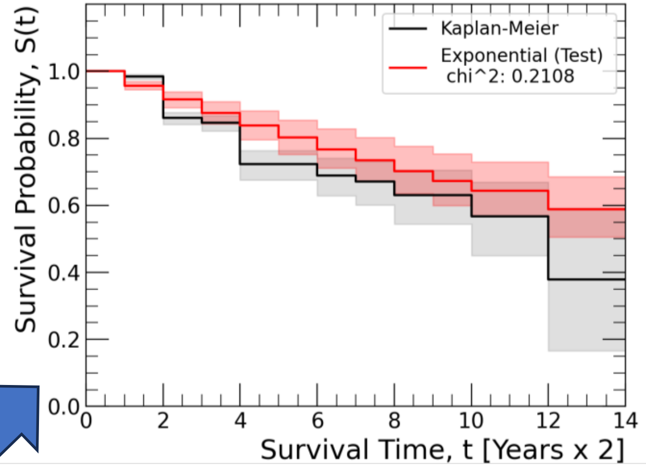
$$S(t_j) = S(t_{j-1})\left(1 - \frac{d_j}{n_j}\right)$$

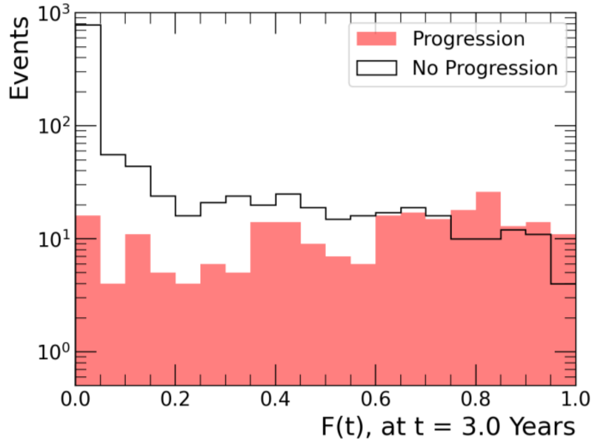
- $S(t_j)$ = prob. of surviving time t_j
- $S(t_{j-1})$ = prob. of surviving at $t_j - 1$
- n_j = No.# of patients alive just before t_j
- d_j = No.# of events at t_j



- What the model is learning:
  - $S(t) = e^{-\lambda t e^{x\beta}}$
- Additional term, $e^{x\beta}$, adds to the function:
  - x : Vector containing image and time information
  - β: Shape parameter that the model learns





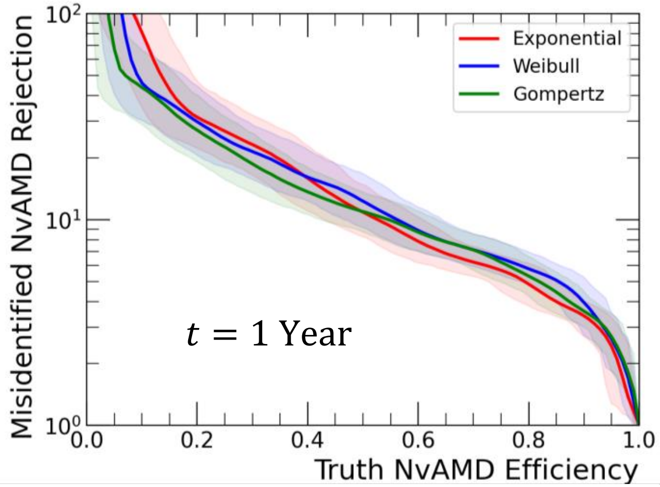


Survival Models – ROC Curves

With 95% Confidence Intervals

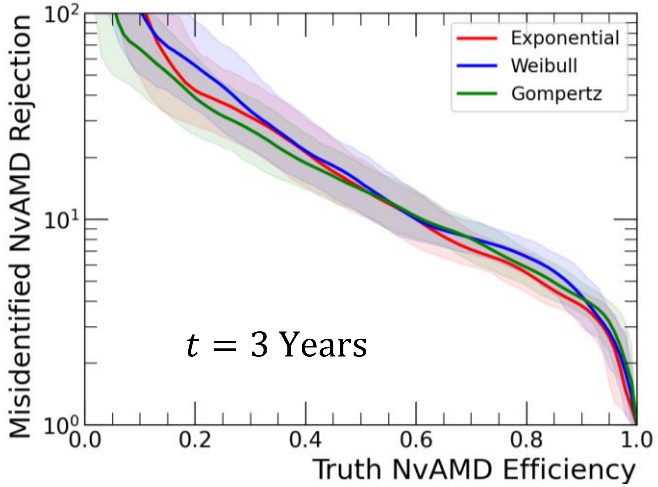
- Efficiency (TPR) $= \frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}} (\times 100\%)$
- Rejection $\left(\frac{1}{\text{FPR}}\right) = \frac{\text{False Positives}+\text{True Negatives}}{\text{False Positives}}$

	Rejection (at Efficiency)		
	0.4	0.6	0.8
Exponential	16	8	4.9
Weibull	16	9	5.8
Gompertz	14	9	5.2



- As time progresses, TP's increase
  - Agrees with model assumption
- Weibull maintains higher rejection at each efficiency (except for $t = 3$ Years at 0.6 efficiency)
- However, rejection is close between models at these points
  - Other metrics required for model choice decision

	Rejection (at Efficiency)		
	0.4	0.6	0.8
Exponential	21	10	5.5
Weibull	21	10	6.5
Gompertz	19	11	5.9



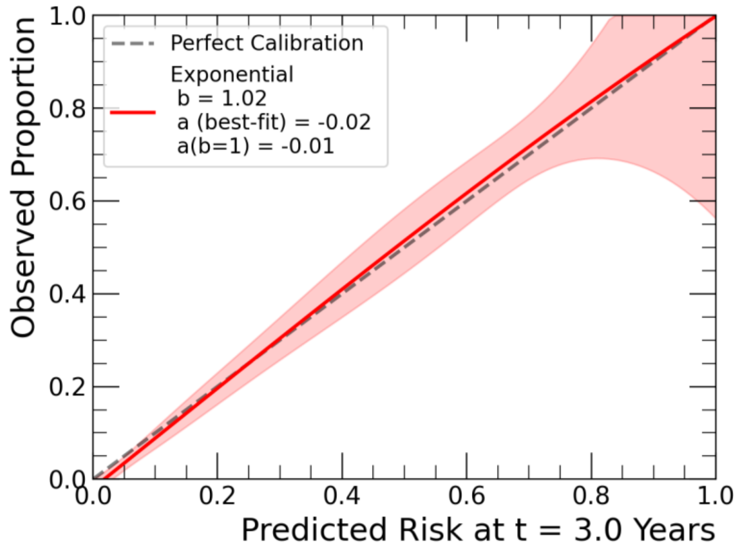
Survival Models - Calibration and Decision Curves

Examples

Calibration Curves

Example – Exponential Baseline Hazard

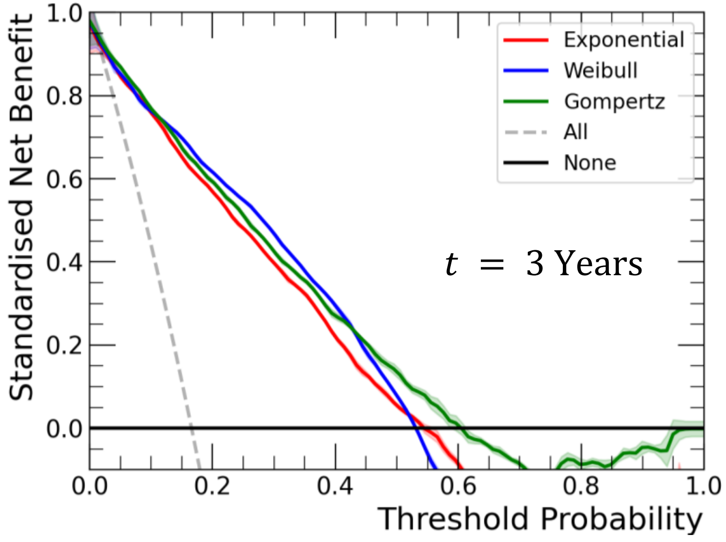
- Shows how well predicted probabilities match actual outcomes
  - ideally, predictions should align with the observed frequencies along the diagonal



Decision Curves

- Net Benefit balances TPs against FPs as a measure of the usefulness of a model, considering the relative harm of FPs

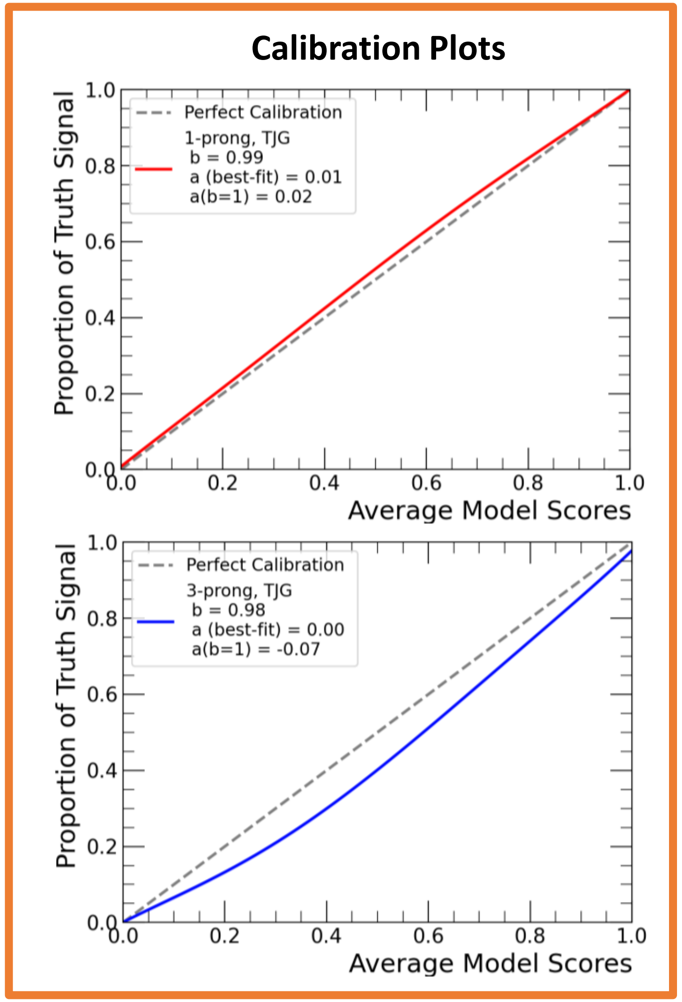
$$\text{Net Benefit} = \frac{\text{True Positives}}{n} - \frac{\text{False Positives}}{n}\left(\frac{p_t}{1-p_t}\right)$$

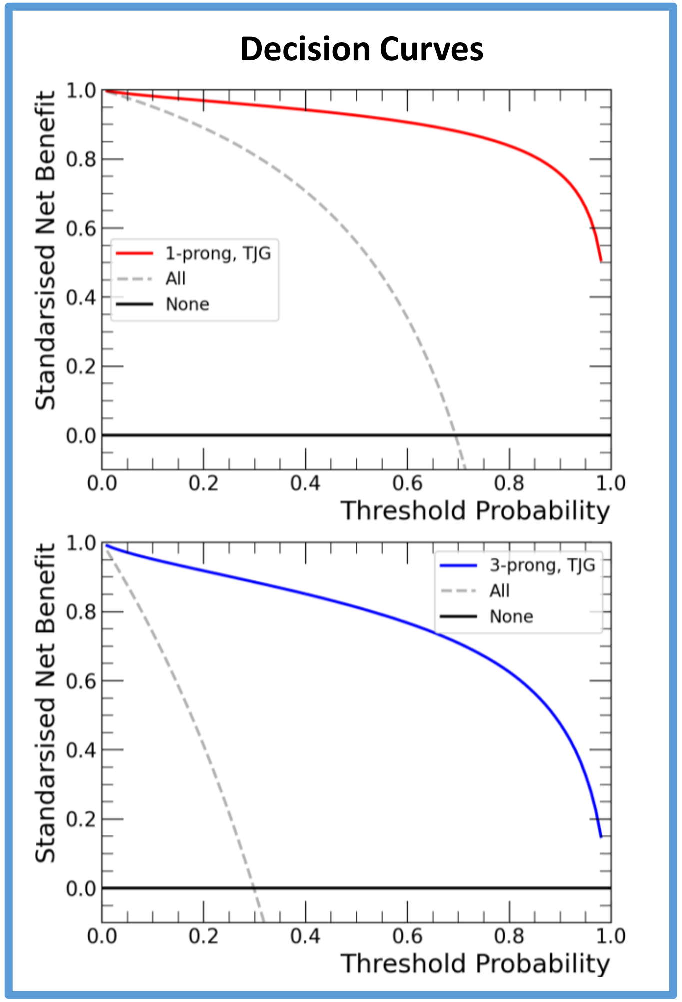


TauJetGraphs – Implementing Medical Analysis Methods

N.B.: Plots made using uncalibrated model scores

- Applying calibration and decision curve analysis to TauJetGraphs can provide useful insights, e.g.:
  - Calibration plots (left) – for 1-prong (top), model predictions reflect likelihood ratios well
  - Decision curves (right) – for the 3-prong (bottom), model shows good benefit in comparison to risks (comparing TPs to FPs w.r.t. a score threshold)
- Decision curves serve as a useful metric for comparing different models (when evaluated on the same datasets)





Summary and Next Steps

Summary:

- GNNs are useful for showing a more natural representation of particle physics data than alternatives, such as RNNs
- Despite being different fields, problems are very similar at their core (classification)
  - There is benefit in using knowledge from one field applied to another!

Next Steps:

- To develop a GNN Survival Model and compare with CNN approach
- To return to physics analysis and utilise knowledge from medical applications of machine learning

Backup

Tau-Leptons at ATLAS

- Hadronic-τ decays, τ_{had}, have a higher branching ratio than the leptonic, τ_{lep} (65% and 35%)
- τ_{had} decays are 1- or 3-prong (1 or 3 $\pi^{\pm}$'s, and maybe some π^{0}'s)
- Problem: τ_{had} signatures are 'drowned-out' by QCD jets

τ_{had} Decay:

- Highly collimated – narrow cone
- Small cross-section
- Low multiplicity

Hadronic Calorimeter
Obtains $\pi^{\pm}$ information

Electromagnetic Calorimeter
Obtains π^{0} information
(via $\pi^{0} \rightarrow \gamma\gamma$ and $\gamma \rightarrow e^{-}e^{+}$)

Tracking Detector
Collects charged particle track information, e.g., direction and position of $\pi^{\pm}$'s from τ-decay

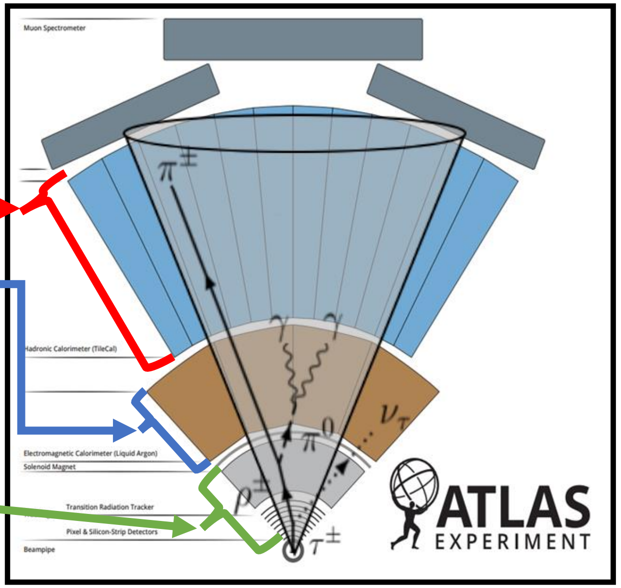


Table: Various τ_{had} decay modes for 1- and 3-prong decays

Hadronic Decay Modes	Label	Branching Ratio, %
$\tau^{\pm} \rightarrow \pi^{\pm}\upsilon_{\tau}$	1p0n	(11.51 ± 0.05)
$\tau^{\pm} \rightarrow \pi^{\pm}\upsilon_{\tau}\pi^{0}$	1p1n	(29.93 ± 0.09)
$\tau^{\pm} \rightarrow \pi^{\pm}\upsilon_{\tau} \geq 2\pi^{0}$	1pXn	(10.81 ± 0.09)
$\tau^{\pm} \rightarrow \pi^{\pm}\pi^{\mp}\pi^{\pm}\upsilon_{\tau}$	3p0n	(9.46 ± 0.05)
$\tau^{\pm} \rightarrow \pi^{\pm}\pi^{\mp}\pi^{\pm}\upsilon_{\tau} \geq \pi^{0}$	3pXn	(5.09 ± 0.05)

Age-Related Eye Disease Study Dataset

- Data is from the (publicly available) Age-Related Eye Disease Study (AREDS) dataset [NCBI]
- Each eye has up to four retinal images taken over time
  - Each image has a time-stamp and whether progression is observed or not
- Clinical information is also available for each patient:
  - Age (at enrolment)
  - Sex (M or F)
  - Smoked (Yes or No)
  - BMI

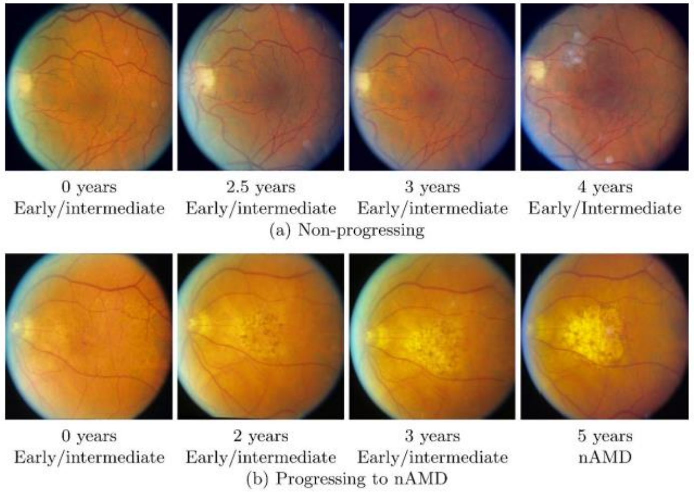


Figure: Example longitudinal images of AMD. Each row contains images taken from the same eye of a patient over time. The first three images in each row displays early or intermediate AMD. The fourth image for patient a) shows that the patient hasn't progressed, while for b) the patient has progressed to an advanced form of AMD. Images are from the AREDS dataset. Figure available from [J. Bridge, 2022, PhD Thesis]

Table: Statistical information on the portion of the AREDS dataset used in this work. Information taken from [J. Bridge, 2022, PhD Thesis]

	Training	Validation	Testing
Eyes	2785	1392	1392
Patients	1532	755	754
Female (%)	1528 (54.9%)	782 (56.2%)	794 (57%)
Mean Baseline Age (range)	74.4 (58.4, 87.9)	74.4 (56.9, 85.5)	74.7 (56.9, 87.8)
Mean Follow-Up Years (Range)	1.3 (0.5, 8.0)	1.3 (0.5, 12.0)	1.24 (0.5, 6.0)
Progressing (%)	476 (17.1%)	238 (17.1%)	238 (17.1%)
Mean BMI at Baseline (Range)	27.5 (8.9, 58.2)	27.4 (15.5, 54.9)	27.2 (16.1, 47.1)
Ever Smoked (%)	1499 (53.8%)	755 (55.7%)	689 (49.5%)

Age-Related Eye Disease Study Dataset

Signal Stats for Each Year, t

	Training Dataset (Signal)		Validation Dataset (Signal)		Testing Dataset (Signal)	
	Count	%	Count	%	Count	%
1 Year	317	11.4	157	11.3	192	13.8
2 Years	449	16.1	225	16.2	228	16.4
3 Years	460	16.5	232	16.7	231	16.6
Prevalence* (Total Signal)	476	17.1	238	17.1	238	17.1

*Prevalence is the rate of disease progression from the dataset:

- Event labels determined from final observation (i.e. 1 if progression observed and 0 if censored)
- Calculated as percentage of progressing eyes from total dataset

Censored Data

Table: Definitions of each type of censored data; left, right, and interval.

Type of Censoring	Definition
Left	Event already happened before patient enrolled into study
Right	Patient leaves study before event is observed
Interval	Event occurs between two observations and exact time is unclear

- AREDS Dataset contains right-censored data, meaning that there are patients who were not observed to progress to advanced forms of AMD before they left the study
- 'Censored' events can only be included in analysis up to the time of 'censoring' – i.e. they must then be 'hidden'/removed as they provide no information past this time

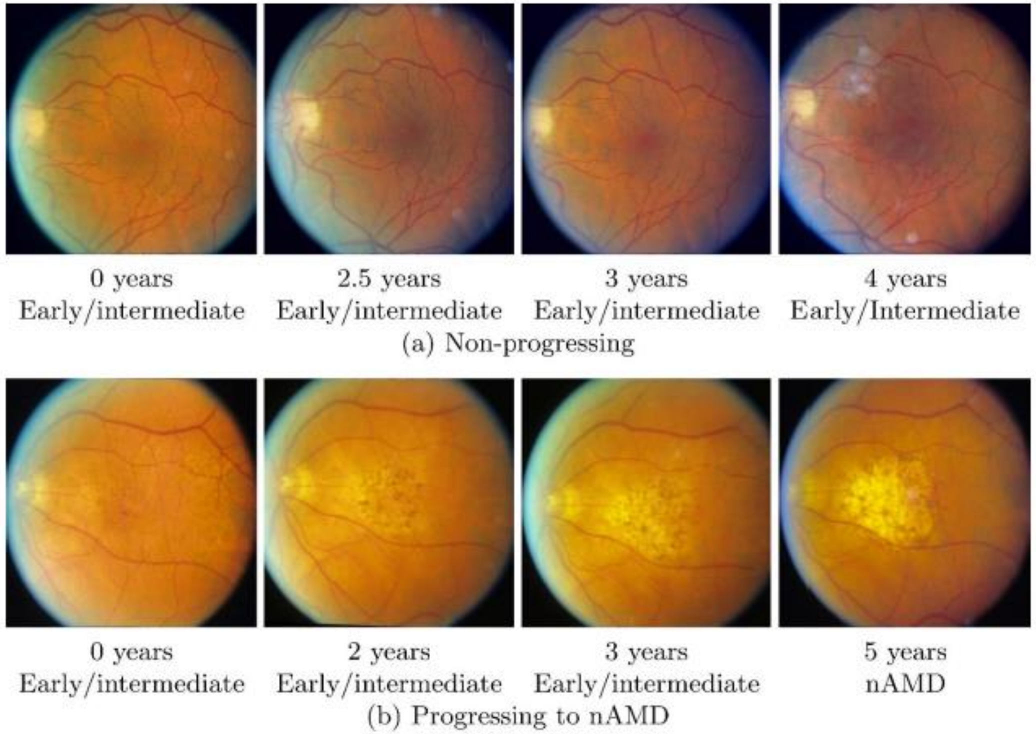


Figure: Example longitudinal images of AMD. Each row contains images taken from the same eye of a patient over time. The first three images in each row displays early or intermediate AMD. The fourth image for patient a) shows that the patient hasn't progressed, while for b) the patient has progressed to an advanced form of AMD. Images are from the AREDS dataset. Figure available from [J. Bridge, 2022, PhD Thesis]

Motivations and Goals

	τ_{had} ID (TauJetGraphs)	AMD Progression (Survival Models)

Motivations	• BR for τ_{had} ($\sim$65%) is $\sim 2\times$ the BR for τ_{lep} ($\sim$35%)
	• ID is important across several research areas, such as:
	  • $H \rightarrow \tau\tau$ production [CERN-EP-2021-217]
	  • Di-Higgs searches with $b\bar{b}\tau^{+}\tau^{-}$
		• AMD is a degenerative disease – it is likely that all patients will progress to an advanced form, given enough time
		• There is no cure, but treatments exist which can help slow the progression
		  • Finding out when a patient is likely to progress is beneficial, as it allows clinicians to appropriately plan treatments and future visits

Goals	• To further study the unification of DMC & ID with a GNN:
	  • Which should be able to handle τ_{had} candidates with 1 and 3 tracks
	  • Final classifier should be able to classify 5 decay modes & QCD jets
		• To modify, extend, and finalise a Time-Distributed CNN, referred to as a Survival Model
		  • The model should be capable of returning a probability of a patient progressing by a given time, t

Survival Models – Mixed-Effects Layer

- Mixed-effects (ME) layer used to model spatial relationships
- The mixed-effects for the i^{th} eye is given by:

$$X_i = \begin{bmatrix} F_{1,1} & \cdots & F_{1,2049} \\ F_{2,1} & \cdots & F_{2,2049} \\ F_{3,1} & \cdots & F_{3,2049} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{2049} \end{bmatrix} + \begin{bmatrix} 1 & 0 & \frac{1}{t_0 - t_1} & \frac{1}{t_0 - t_2} \\ 1 & \frac{1}{t_1 - t_0} & 0 & \frac{1}{t_1 - t_2} \\ 1 & \frac{1}{t_2 - t_0} & \frac{1}{t_2 - t_1} & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + [\epsilon_1 \quad \epsilon_2 \quad \epsilon_3]$$

Where:

- F_1, F_2, and F_3 are feature vectors extracted by CNN into fixed-effects matrix
- α are fixed-effects parameters (learned by the model)
- t_0, t_1, and t_2 are observation times, with t_0 being initial observation (random-effects matrix, Z)
- β are random-effects parameters (learned by the model)
- ϵ are unknown random errors
- ME layer results in a single vector, X_i, with relationships between time points modelled using Z

Terms:

1. **Fixed-effects –** models the relationship within slices/images
2. **Random-effects –** models the spatial relationship between slices/images

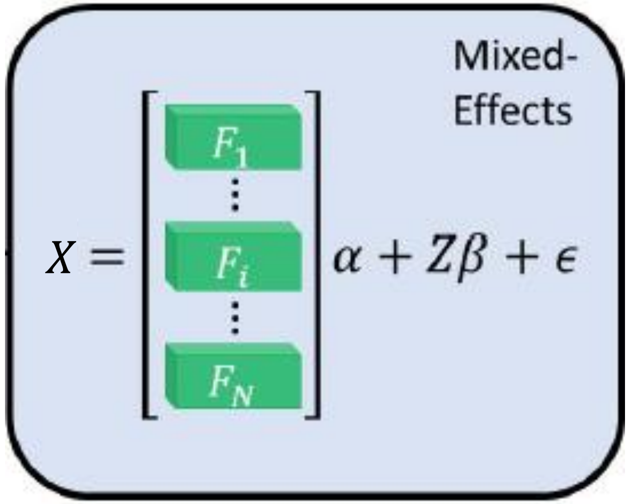


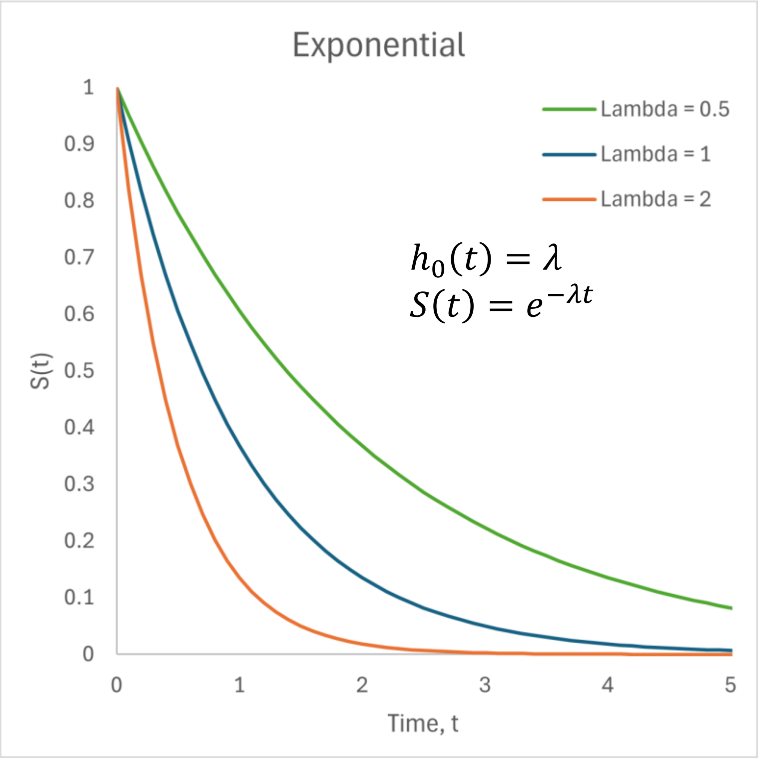
Figure: The mixed-effects layer of the model architecture. Figure available from [J. Bridge, 2022, PhD Thesis]

Survival Models & Functions

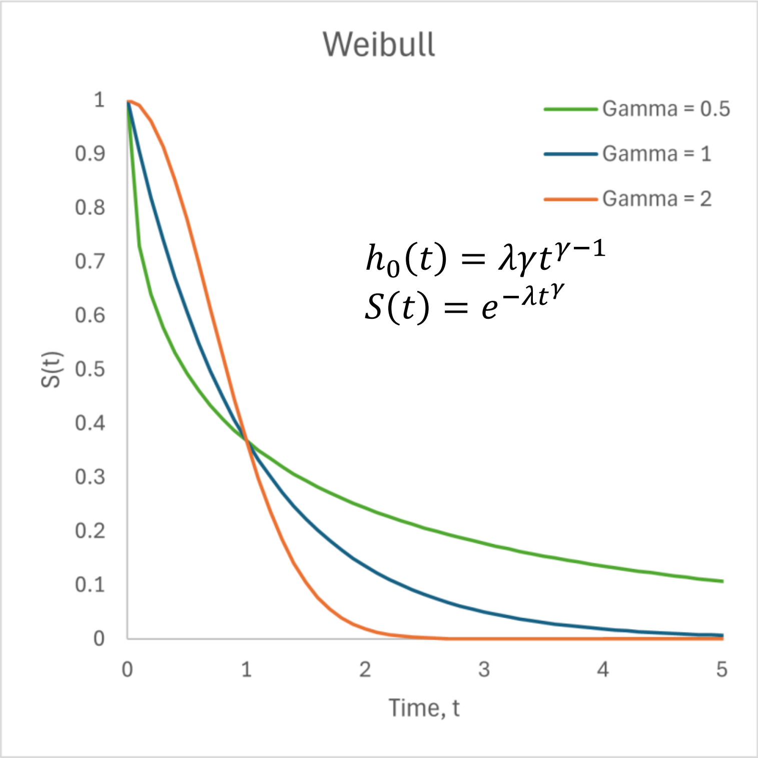
- Survival models estimate the probability of an event occurring up to a specified time, modelled by a survival function, $S(t)$
- Probability of event at time, t, determined by hazard function, $h(t)$
  - Where: $h_0(t)$ is the baseline hazard function; x the resulting ME vector, X_i; and β a vector of parameters
- Three models trained – each using a different baseline hazard function, $h_0(t)$:
  1. Exponential – $h_0(t) = \lambda \rightarrow S(t) = e^{-\lambda t e^{x\beta}}$
  2. Weibull – $h_0(t) = \lambda\gamma t^{\gamma-1} \rightarrow S(t) = e^{-\lambda t^{\gamma} e^{x\beta}}$
  3. Gompertz – $h_0(t) = \lambda e^{-\gamma t} \rightarrow S(t) = e^{-\left(\frac{\lambda}{\gamma}\right)(e^{\gamma t}-1)e^{x\beta}}$

Name	Equation
Hazard function	$h(t) = h_0(t)e^{x\beta}$
Cumulative hazard function	$H(t) = \int_0^t h(u)\,du$
Survival function	$S(t) = P(T \geq t) = \exp\{-H(t)\}$
Prob. Of Failure by time, t (Cumulative Density Function)	$F(t) = P(T < t) = 1 - S(t)$
Probability Density Function	$f(t) = \frac{d}{dt}F(t) = h(t)S(t)$

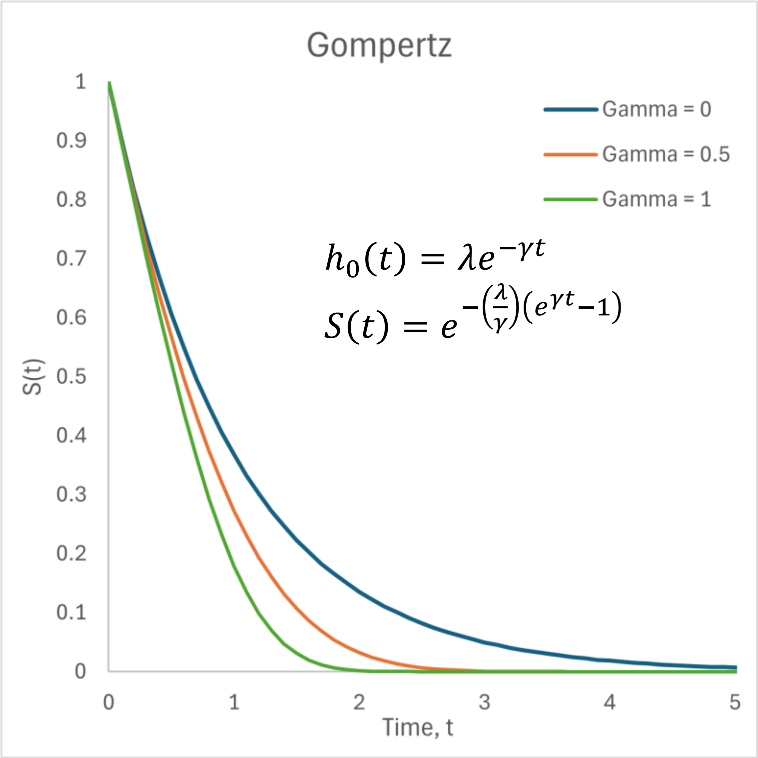
Survival Function For Each Baseline Hazard



- Assumes constant hazard rate → λ = constant
- Higher λ = higher hazard rate → lower survival

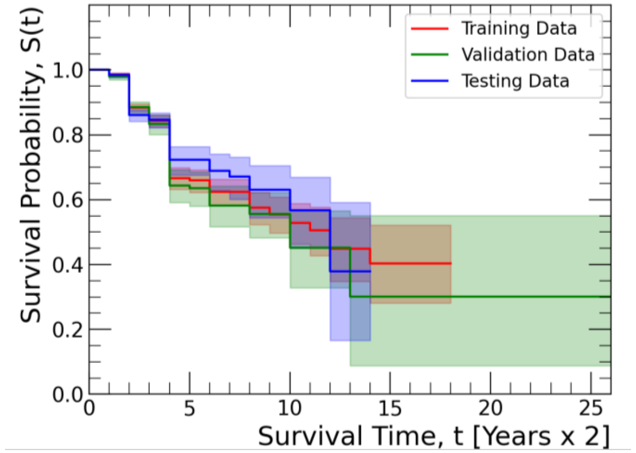


- Plotted with λ =1
- Matches exponential distribution when γ = 1



- Plotted with λ =1
- Matches exponential distribution when γ = 0

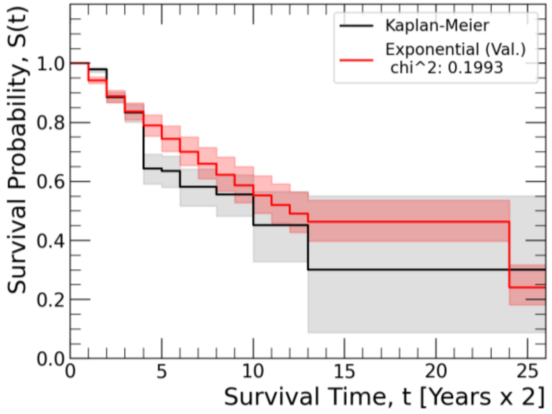
Survival Models – Kaplan-Meier Curves

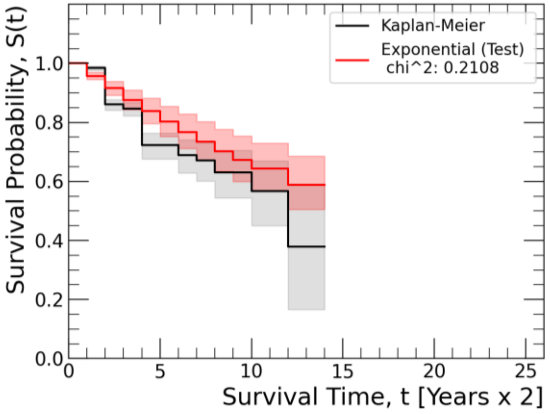


Validation Dataset:

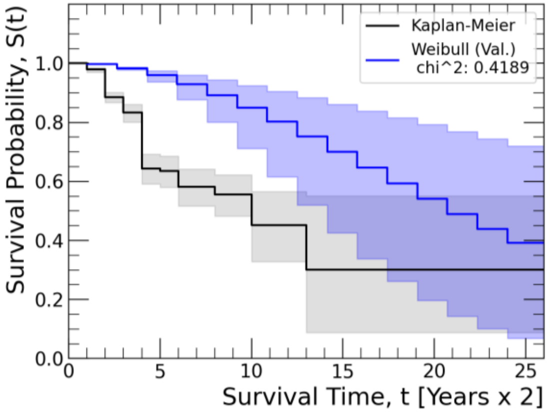
Testing Dataset:

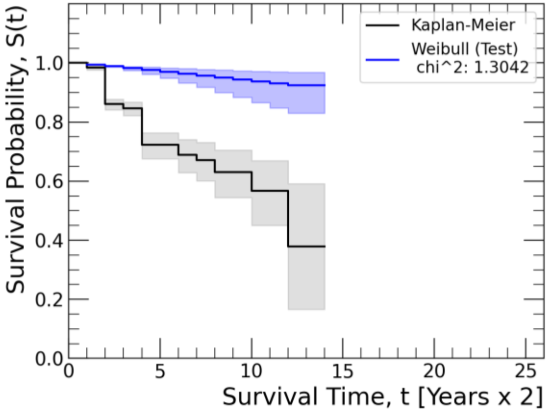
Baseline: Exponential



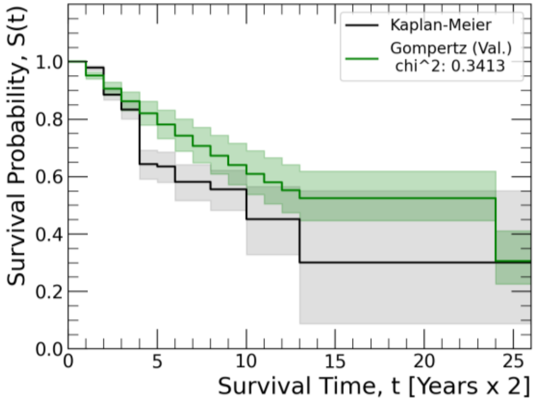


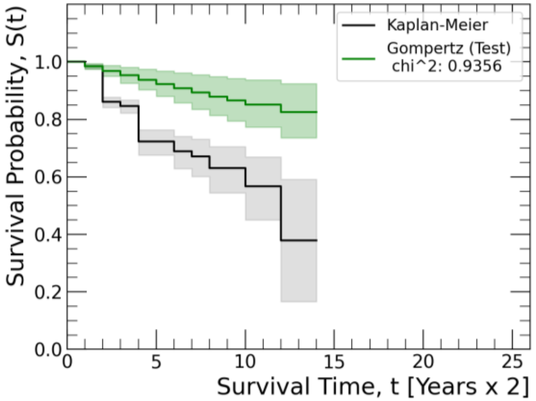
Baseline: Weibull



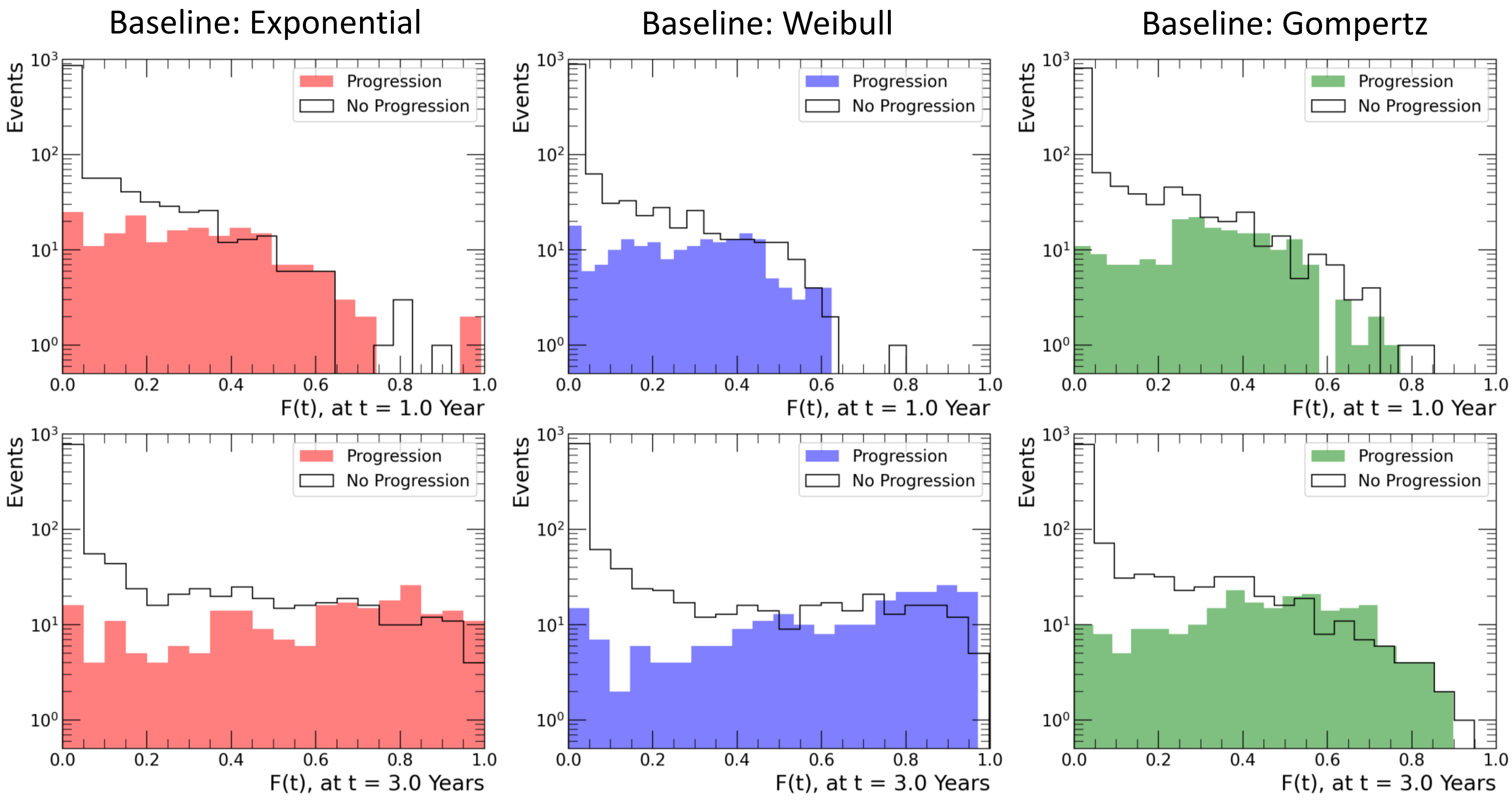


Baseline: Gompertz





Survival Models - Output Score (Risk) Distributions



Decision Curves – Background

- Net Benefit calculated using:

$$\text{Net Benefit} = \frac{\text{TP}}{\text{n}} - \left(\frac{\text{FP}}{\text{n}} \times \frac{\text{p}_\text{t}}{1 - \text{p}_\text{t}}\right)$$

Where TP = # True Positives, FP = # False Positives, n = Number of events, p_t = score threshold

An example using the figure:

- A model exists that provides a probability of a patient has a disease:
  - If near 1, model is confident they have the disease and they will ask to be treated, and similarly if it is near 0 then the model is confident that they don't have the disease and so won't ask to be treated
  - There exists a probability between 0 and 1 where the patient is unsure whether they will forgo treatment
    - This threshold probability, p_t, is where the benefit of treatment is equal to the expected benefit of avoiding treatment.
  - Solving this from the figure => $p_t a + (1 - p_t)b = p_t c + (1 - p_t)d$
  - Becoming: $\frac{a-c}{d-b} = \frac{1-p_t}{p_t}$, where *d-b* is the consequence of being treated unnecessarily (harm associated with FP result), and *a-c* is the consequence of avoiding treatment when it would have been of benefit (harm from FN result)
  - "Harm" is considered as the overall effect of negative consequences of a particular decision

For p = probability of disease, and a, b, c, and d give the value associated with each outcome in terms such as quality-adjusted life-years. Figure from [doi: 10.1177/0272989X06295361]

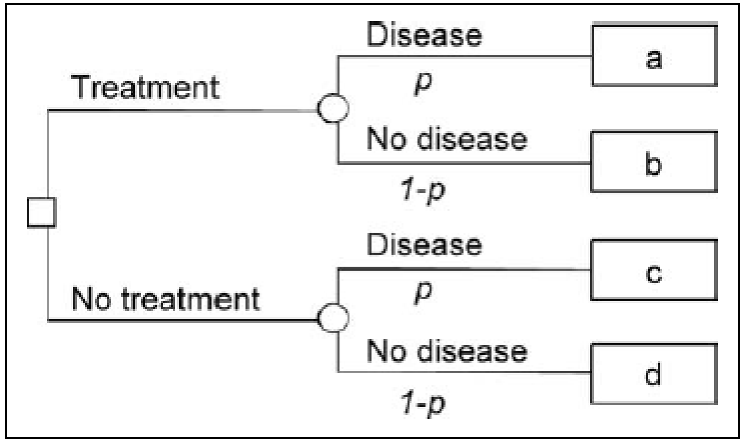


Figure 1 A decision tree for treatment. The probability of disease is given by p; a, b, c, and d give, respectively, the value of true positive, false positive, false negative, and true negative.

Inputs used for TauID RNN, ATL-PHYS-PUB-2019-033,
also used in TauJetGraphs GNN

	Observable	1-prong	3-prong
Track inputs	$p_{\mathrm{T}}^{\text{seed jet}}$	●	●
	$p_{\mathrm{T}}^{\text{track}}$	●	●
	$\Delta\eta^{\text{track}}$	●	●
	$\Delta\phi^{\text{track}}$	●	●
	$\lvert d_0^{\text{track}}\rvert$	●	●
	$\lvert z_0^{\text{track}} \sin\theta\rvert$	●	●
	$N_{\text{IBL hits}}$	●	●
	$N_{\text{Pixel hits}}$	●	●
	$N_{\text{SCT hits}}$	●	●

	Observable	1-prong	3-prong
Cluster inputs	$p_{\mathrm{T}}^{\text{jet seed}}$	●	●
	$E_{\mathrm{T}}^{\text{cluster}}$	●	●
	$\Delta\eta^{\text{cluster}}$	●	●
	$\Delta\phi^{\text{cluster}}$	●	●
	λ_{cluster}	●	●
	$\langle\lambda^2_{\text{cluster}}\rangle$	●	●
	$\langle r^2_{\text{cluster}}\rangle$	●	●

	Observable	1-prong	3-prong
High-level inputs	$p_{\mathrm{T}}^{\text{uncalibrated}}$	●	●
	f_{cent}	●	●
	$f^{-1}_{\text{leadtrack}}$	●	●
	ΔR_{max}	●	●
	$\lvert S_{\text{leadtrack}}\rvert$	●	
	$S_{\mathrm{T}}^{\text{flight}}$		●
	$f_{\text{iso}}^{\text{track}}$	●	●
	$f_{\text{track}}^{\text{EM}}$	●	●
	$p_{\mathrm{T}}^{\text{EM+track}}/p_{\mathrm{T}}$	●	●
	$m^{\text{EM+track}}$	●	●
	m^{track}		●

Variable	Description
$p_T(\tau_{had})$	p_T of the τ_{had} (using calorimeter based $\tau_{had\text{-}vis}$ energy scale)
$p_T(\text{object})$	p_T of the object
$\Delta\phi(\text{object}, \tau_{had})$	Distance between the object and τ_{had} in ϕ
$\Delta\eta(\text{object}, \tau_{had})$	Distance between the object and τ_{had} in η
$\Delta\phi(\text{object, trackECal})$	Distance between the object and the extrapolation of highest-p_T τ_{had} track to EM calorimeter in ϕ
$\Delta\eta(\text{object, trackECal})$	Distance between the object and the extrapolation of highest-p_T τ_{had} track to EM calorimeter in η

Physics object kinematic variables

Variables used in Decay Mode Classification DSNN, ATL-PHYS-PUB-2022-044, also utilised in TauJetGraphs GNN

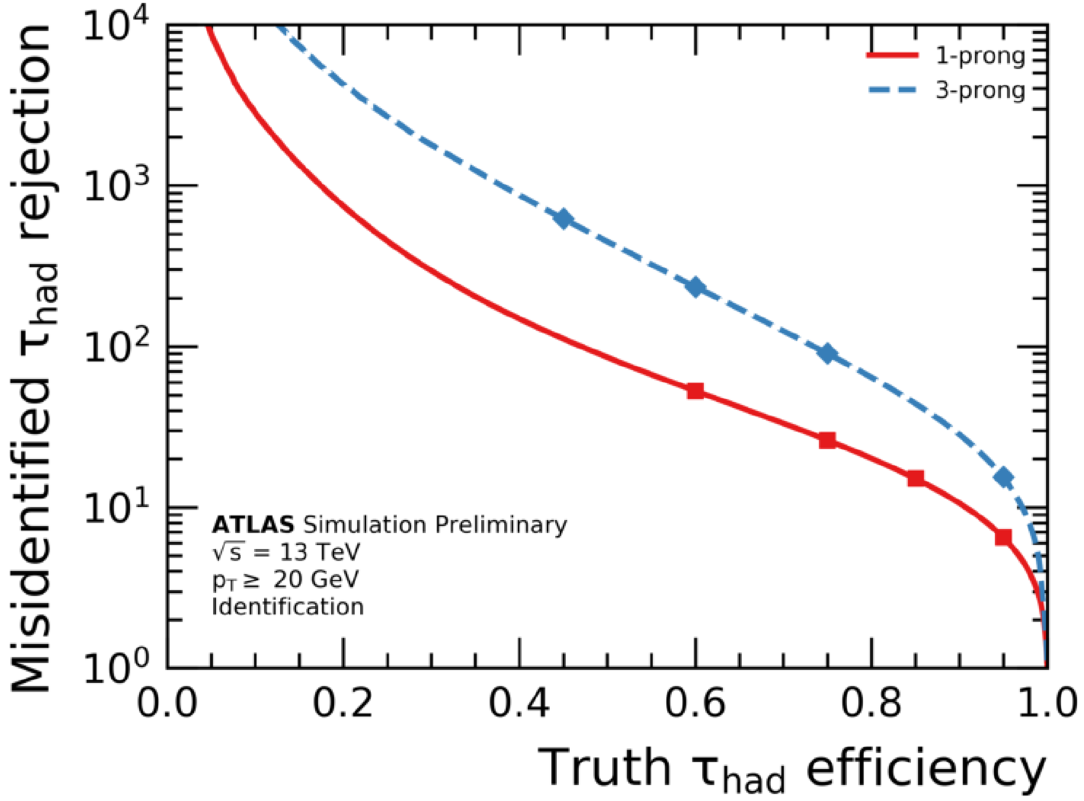
$\langle\eta^1\rangle$	First moment in η in cluster shower axis
$\log(\langle r^2\rangle)$	Second moment in the radial distance of cluster cells from the shower axis
$\Delta\theta$	Distance in θ between the EM shower axis and the vector pointing from the primary vertex to the centre of the shower
$\log(\lambda_{centre})$	Distance of the cluster shower centre from the calorimeter front face measured along the shower axis
$\langle\lambda^2\rangle$	Mean distance of a cell from the shower centre along the shower axis
$\log(\langle\rho^2\rangle)$	Second moment in the cluster energy density, where $\rho = E^{\text{cluster}}/V^{\text{cluster}}$
f_{core}	Sum of energy fractions in the most energetic cells per sampling
f_{core}^{EM1}	Same as f_{core} but only consider EM1
$N_{pos,EM1}$	Number of cells with positive energy in EM1
$N_{pos,EM2}$	Number of cells with positive energy in EM2
E_{EM1}	Energy in the EM1 layer
E_{EM2}	Energy in the EM2 layer
$\langle\eta^1_{EM1}\rangle$ w.r.t. cluster	First moment in η in EM1 with respect to the cluster
$\langle\eta^1_{EM2}\rangle$ w.r.t. cluster	First moment in η in EM2 with respect to the cluster
$\log(\langle\eta^2_{EM1}\rangle)$ w.r.t. cluster	Second moment in η in EM1 with respect to the cluster
$\log(\langle\eta^2_{EM2}\rangle)$ w.r.t. cluster	Second moment in η in EM2 with respect to the cluster

Neutral pion cluster variables

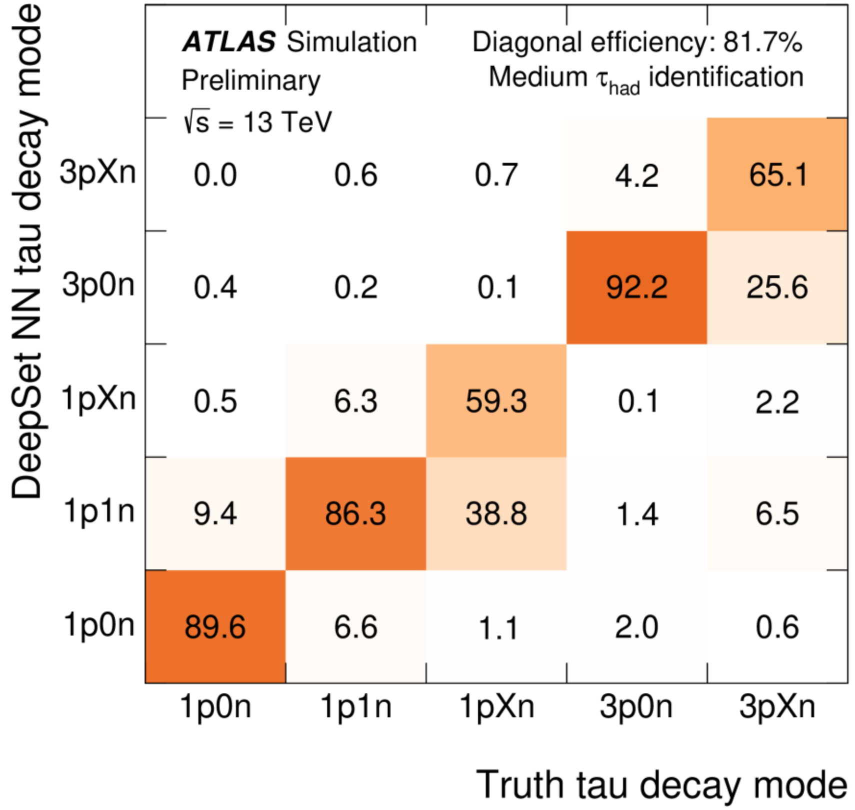
RNN – ROC Curve and Confusion Matrix

Plots taken from: [ATL-PHYS-PUB-2022-044]

Identification ROC Curve



Decay Mode Classification Confusion Matrix



DeepSet NN \ Truth	1p0n	1p1n	1pXn	3p0n	3pXn
3pXn	0.0	0.6	0.7	4.2	65.1
3p0n	0.4	0.2	0.1	92.2	25.6
1pXn	0.5	6.3	59.3	0.1	2.2
1p1n	9.4	86.3	38.8	1.4	6.5
1p0n	89.6	6.6	1.1	2.0	0.6

Metric Definitions

- **Accuracy** – The fraction of correctly classified samples (if normalised = True)
- **Purity (Precision)** – Purity is the measure of how well a classifier avoids incorrectly labelling a sample as positive. It's calculated as true positives divided by true positives plus false positives:
  - $\frac{tp}{tp+fp}$ where *tp* is true positive and *fp* is false positive
- **Efficiency (Recall)** – Efficiency measures how well a classifier finds all the true positives. It's calculated as true positives divided by true positives plus false negatives:
  - $\frac{tp}{tp+fn}$ where *tp* is true positive and *fn* is false negative
- **Background Rejection –** The inverse of the Background Selection Efficiency, depending on the Signal Selection Efficiency

Glossary

- **AMD** – Age-related Macular Degeneration
- **AREDS** – Age-Related Eye Disease Study
- **ID –** Identification
- **DMC –** Decay Mode Classification
- $\boldsymbol{\tau}_{\mathbf{had}}$ **-** Hadronically decaying τ-lepton
- **RNN –** Recurrent Neural Network
- **DSNN -** DeepSet Neural Network
- **GNN –** Graph Neural Network
- **CNN** – Convolutional Neural Network
- **ROC Curve -** Receiver Operator Characteristic Curve