UNIWERSYTET JAGIELLOŃSKI W KRAKOWIE

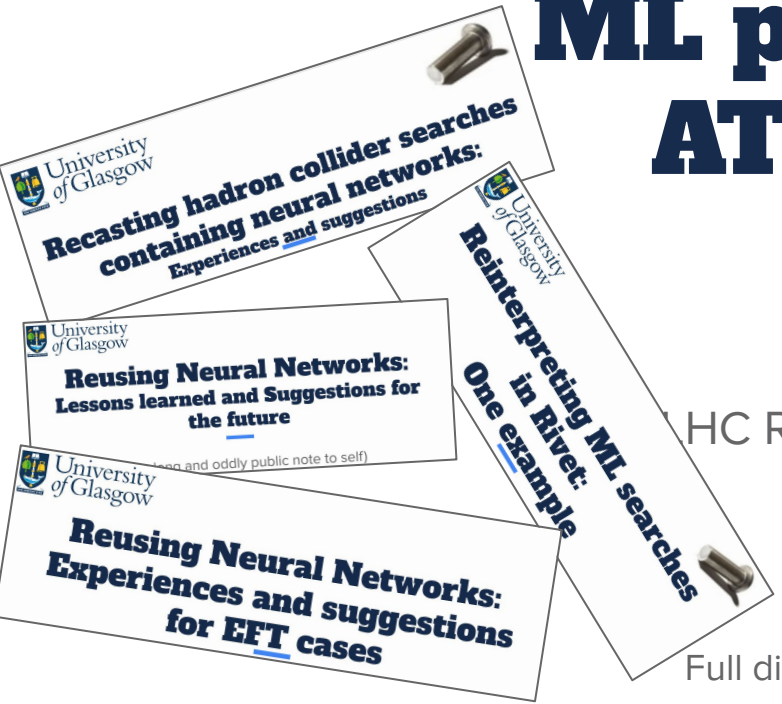
ATLAS EXPERIMENT

ML preservation: ATLAS' view

Tomasz Procter,
LHC Reinterpretation Forum,
February 2025

Grant CHIST-ERA 2022/04/Y/ST2/00186 National Science Centre, Poland

ML preservation: ATLAS' view



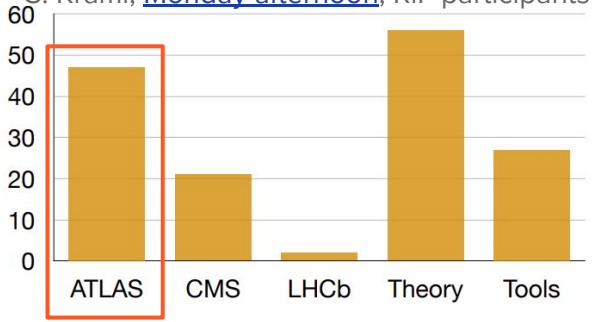
Tomasz Procter,
LHC Reinterpretation Forum,
February 2025

Full disclosure: I've given this talk from the other side...

ML preservation: ATLAS' view

Tomasz Procter,
LHC Reinterpretation Forum,
February 2025

S. Kraml, Monday afternoon, RiF participants



~1% of the collaboration is here: ~50 participants => 50 ATLAS views?

Outline

- The past: what has ATLAS done so far
  - Analyses with BDTs
  - Analyses with NNs
  - Why so few?
  - General comments
- The future: problems and opportunities going forward:
  - Lower-level inputs
  - GN2

Published BDTs

Analysis	BDT Format	Location
ATLAS-SUSY-2016-16 (top-squark par prod. with one lep, jets, & MET 36fb^{-1})	xml files	HepData & SimpleAnalysis
ATLAS-SUSY-2018-22 (squarks and gluinos, with jets & MET, 139fb^{-1})	BDT as C++ code (petrify-BDT)	HepData & SimpleAnalysis
ATLAS-SUSY-2019-02 (pair prod. of sleptons & charginos decaying to 2 lep & neutralinos with mass splittings near the W mass, 139fb^{-1}	ROOT BDT	HepData & SimpleAnalysis
ATLAS-SUSY-2020-16 (Searches for EW prod. of sparticles with compressed mass spectra, 139 fb^{-1})	ROOT BDT	SimpleAnalysis
ATLAS-SUSY-2023-26 (vector boson fusion signatures and missing transverse momentum, 139 fb^{-1})	ROOT BDT	HepData & SimpleAnalysis
ATLAS-EXOT-2022-04 (Search for neutral LLPs decay into displaced jets in the calorimeter in association with leps or jets 139^{-1})	Files+python examples	HepData

See talk this morning!

Published BDTs

Analysis	BDT Format	Location
ATLAS-SUSY-2016-16 (top-squark par prod. with one lep, jets, & MET 36fb^{-1})	xml files	HepData & SimpleAnalysis
ATLAS-SUSY-2018-22 (squarks and gluinos, with jets & MET, 139fb^{-1})		& SimpleAnalysis
ATLAS-SUSY-2019-02 (pair prod. of sleptons & charginos decaying to splittings near the W mass, 139fb^{-1}		& SimpleAnalysis
ATLAS-SUSY-2020-16 (Searches for EW prod. of sparticles with comp		lysis
ATLAS-SUSY-2023-26 (vector boson fusion signatures and missing transverse momentum, 139 fb^{-1})		& SimpleAnalysis
ATLAS-EXOT-2022-24 (Search for neutral LLPs decay into displaced jets in the calorimeter in association with leps or jets 139^{-1})	Files+python examples	HepData

Received broadly positive feedback from reinterpreters

See talk this morning!

Published BDTs

Analysis	BDT Format	Location
ATLAS-SUSY-2016-16 (top-squark par prod. with one lep, jets, & MET 36fb^{-1})	xml files	HepData & SimpleAnalysis
ATLAS-SUSY-2018-22 (squarks and gluinos, with jets & MET, 139fb^{-1})	BDT as C++ code	HepData & SimpleAnalysis
ATLAS-SUSY-2019-02 (pair prod. of sleptons & charginos decaying to 2 lep & neutralinos with mass splittings near the W mass, 139fb^{-1}	ROOT BDT	HepData & SimpleAnalysis
ATLAS-SUSY-2020-16 (Searches for EW prod. of sparticles with compressed mass spectra, 139		
ATLAS-SUSY-2023-26 (vector boson fusion signatures and missing transverse momentum, 139		is
ATLAS-EXOT-2022-24 (Search for neutral LLPs decay into displaced jets in the calorimeter in association with leps or jets 139^{-1})		

Used in (published) pMSSM scan paper (both simpleAnalysis and RECAST)

talk this ning!

Published Neural Nets

Analysis	**NN Type**	**NN format**	**Location**
ATLAS-SUSY-2019-04 (R-parity violating SUSY with leps and many jets)	S vs B classifier DNN	ONNX	HepData & SimpleAnalysis
ATLAS-SUSY-2019-04 (SUSY MET + multi-bjets)	S vs B classifier DNN	ONNX	SimpleAnalysis
ATLAS-EXOT-2019-23* (neutral LLPs into displaced hadronic jets) *Also contained & published BDTs	S vs B BDT (event-level) S vs B NN (per-jet)	ONNX (low-level); **Efficiency maps** (high-level)	HepData
ATLAS-HDBS-2019-23 (Anomaly detection search for resonances decaying to Higgs+X)	VRNN for anomaly detection	Weights file (keras?); python example	HepData
+ More coming soon...			

Published Neural Nets

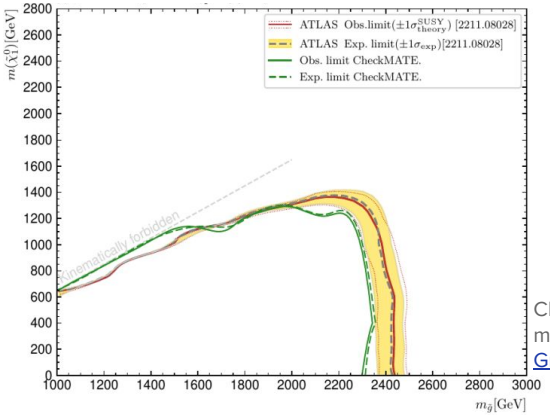
Analysis	NN Type	NN format	Location
ATLAS-SUSY-2019-04 (R-parity violating SUSY with leps and many jets)	S vs B classifier DNN	ONNX	HepData & SimpleAnalysis
ATLAS-SUSY-2019-04 (SUSY MET + multi-bjets)			
ATLAS-EXOT-2019-23* (neutral LLPs into displaced hadronic jets) *Also contained & published BDTs			
ATLAS-HDBS-2019-23 (Anomaly detection search for resonances decaying to Higgs+X)	detection	example	
+ More coming soon...			

Proved difficult to reinterpret
(though this also applied to C&C regions of the analysis too)

Published Neural Nets

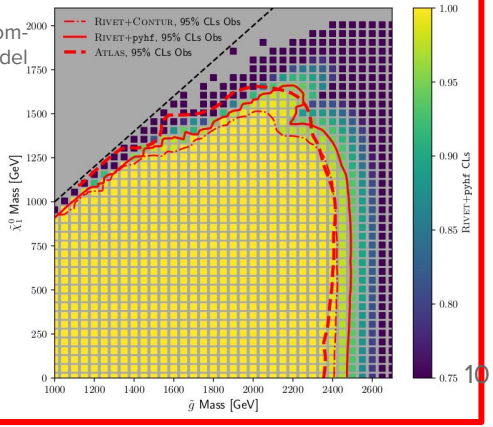
Analysis	NN Type	NN format	Location
ATLAS-SUSY-2019-04 (R-parity violating SUSY with leps and many jets)	S vs B classifier DNN	ONNX	HepData & SimpleAnalysis
ATLAS-SUSY-2019-04 (SUSY MET + multi-bjets)	S vs B classifier DNN	ONNX	SimpleAnalysis
ATLAS-EXOT-2019-23* (neutral LLPs into displaced hadronic jets) *Also contained & published BDTs			
ATLAS-HDBS-2019-23 (Anomaly detection search for resonances decaying			
+ More coming soon			

Much more successful:



CheckMATE, stop-stop model, from K. Rolbiecki, Grenoble, June 2024

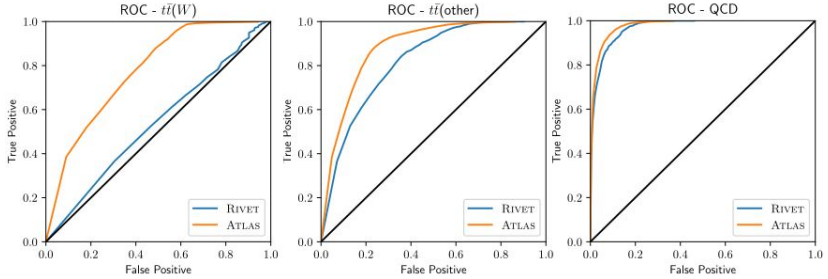
Rivet, sbottom-sbottom model



Published ML - Comments on scarcity

- Publishing weights is not the norm:
  - **7*** published analyses on glance tag LWTNN – **0** published lwtnn files.
  - **32** analyses with “BDT” in a plot or other entry on Hepdata – only **7** come with weights.
- Tends to happen when an analysis team really cares about reinterpretation.
  - Do we need to reconsider rules/procedures on aux. data or similar?
    - What if HEPdata gets delayed?
    - Can we give better credit to particular analysis teams that have gone above and beyond?
  - Again, general reinterpretation problem not unique to (though worse in the context of) ML.
- Run 2: several networks used across many analyses: **0** public.
  - E.g. W-tagger, top-tagger, MCBOT, “belong” to CP groups
    - Some include substructure dependence, reinterpretability is not trivial.

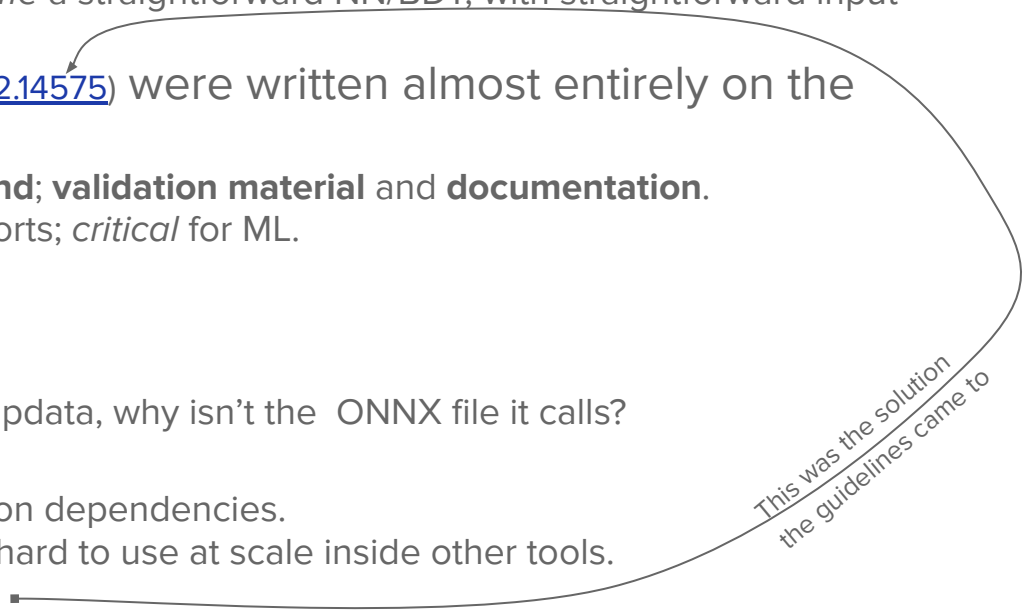
* I know this is an undercount by at least two.



ATLAS Run 2 top-tagger, reinterpreted in Rivet, T. Procter thesis

Published ML - General comments

- ~10 analyses is still LHC-leading!
- Several examples implemented and **validated** in multiple frameworks
  - We're at the stage where we *assume* a straightforward NN/BDT, with straightforward input features, *should* work.
- LH ML-reuse guidelines (arXiv:2312.14575) were written almost entirely on the basis of these ATLAS analyses.
  - Emphasis on: **keeping reuse in mind**; **validation material** and **documentation**.
  - *Important* in all reinterpretation efforts; *critical* for ML.
- Lots of formats/locations:
  - So sorry if I've missed anything!
  - How can we standardise?
  - If the SimpleAnalysis code is on hepdata, why isn't the ONNX file it calls?
- What even is the ideal format?
  - Python code may have many version dependencies.
  - Containers are most "eternal", but hard to use at scale inside other tools.
  - Onnx as the best option (for now?)



Published ML - some questions for ATLAS

- What would the procedure for sharing an ATLAS-wide tool look like?
  - If it comes with the paper/note that introduces the technique, where is the validation?
- SimpleAnalysis codes have been a key mechanism for getting these out.
  - (Nothing especially unique to SimpleAnalysis, any similar framework would probably work given the same (fantastic) institutional support)
  - What's going to be the status of this going forward?
- Can we make it easier to get these on Hepdata?
  - If the SimpleAnalysis code is provided, can the weights be, too?
- How do we make sure that weights are safely stored internally.
  - We all know what can happen when one important person leaves physics/ATLAS.
    - But this can be even scarier in the ML context.
  - I know of at least one (albeit 36 fb^{-1}) analysis where weights **used in the paper** were lost, within a couple of years of publication.
  - Should they all end up in RECAST? Anywhere imbetween?

Going Forward

Run 3 Challenge: low-level inputs and others

- Trend in HEP ML usage is to use more-and-more, lower-and-lower level input features
- Typically require proper detector-sim to get right
  - They may be almost meaningless to those outside the experiment.
- So just publishing the weights alone may not make the analysis re-usable.
- New ML strategies will also make life harder…

What to do?

Run 3 Challenge: low-level inputs and others

- Trend in HEP ML usage is to use more-and-more, lower-and-lower level input features
- Typically require proper detector-sim to get right
  - They may be almost meaningless to those outside the experiment.
- So just publishing the weights alone may not make the analysis re-usable.
- New ML strategies will also make life harder…

What to do?

1. (multi-dimensional) efficiencies?
   - E.g. ATLAS-EXOT-2019-23.
     - 6D parameterised efficiencies.
   - BUT: might not always be appropriate for all BSM classifiers.
   - How many dimensions is enough?

Run 3 Challenge: low-level inputs and others

- Trend in HEP ML usage is to use more-and-more, lower-and-lower level input features
- Typically require proper detector-sim to get right
  - They may be almost meaningless to those outside the experiment.
- So just publishing the weights alone may not make the analysis re-usable.
- New ML strategies will also make life harder…

What to do?

2. Surrogate networks/BDTs:
   - ML trained to replicate the output of a more complex network using truth-level features.
   - In the reinterpretation setting, can “cheat” with extra truth-level information.
   - See talk this morning
     - Once again, LLPs lead the way…
   - Exciting to see where this goes…
   - But one surrogate network per analysis will be a lot of work.

Run 3 Challenge: low-level inputs and others

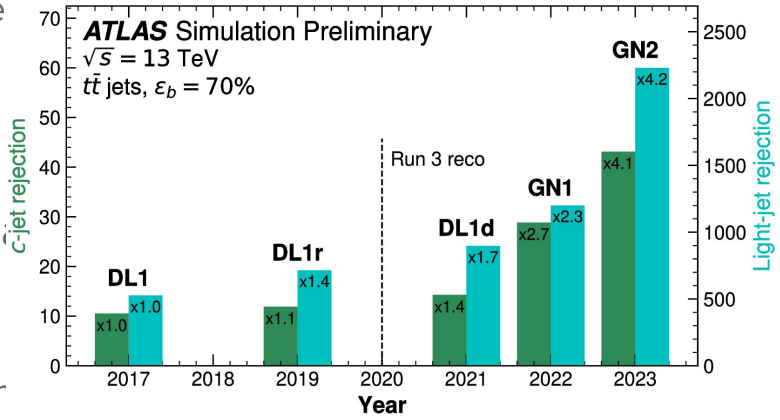
- Trend in HEP ML usage is to use more-and-more, lower-and-lower level input features
- Typically require proper detector-sim to get right
  - They may be almost meaningless to those outside the experiment.
- So just publishing the weights alone may not make the analysis re-usable.
- New ML strategies will also make life harder…

What to do?

3. Does this make RECAST *more* important?
   - For unusual signatures, efficiencies and even surrogates may struggle.
   - The only environment where the network's actual output could be tested.
   - Have we done any tests of ML-based analyses in RECAST?

Run 3: GN2

- ATLAS's "blue ribbon" ML project – factor two performance for FTAG related tasks.
- Should be accompanied by the release of training dataset –
  - With only a few extra truth-level variables, multi-dimensional efficiency maps, surrogates etc. could potentially be "outsourced" to those interested.
- ATLAS-wide, won't place a huge extra burden on small analysis teams.
- This is not far from what most codes alre do with *b*-tagging.
  - But should allow us to do it better.
  - And the wider GN-family of algorithms will do r than just *b*/*c*-tagging.



Conclusion

Conclusions

- ATLAS has published 10+ sets of weights for ML models
  - Provided the material for most of the early testing by the reinterpretation community
  - Several successful examples!

- How do we take the next step from this?
  - Support for analysis teams/procedural simplifications?
  - How are things preserved internally?
  - There will always be some non-resuable networks:
    - How do we deal with networks based on very low-level features?

- GN2 dataset will be very interesting!

BONUS