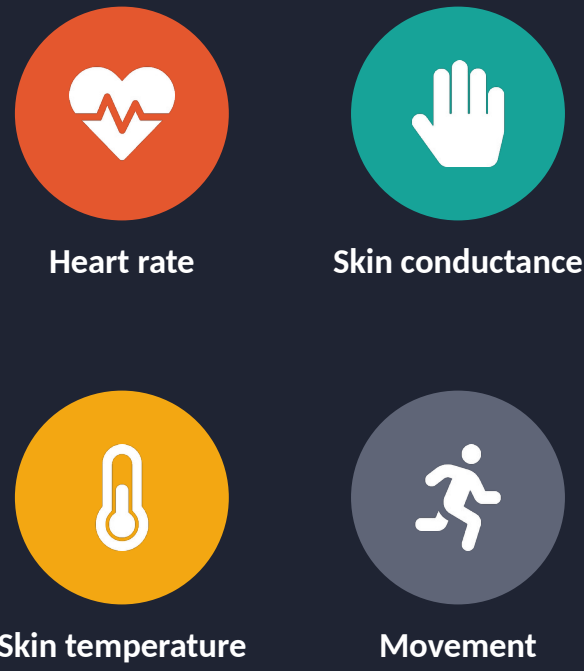
Predicting Autistic Meltdowns from Wearable Biosensor Data

A Comparative Study of Classical and Deep Learning Models



Dhruv Aggarwal

State of North Carolina Undergraduate Research & Creativity Symposium
UNC Charlotte · November 7, 2026

Why predict a meltdown?



Involuntary, and often misread

Meltdowns are a response to overwhelming sensory or emotional input, but they are often mistaken for tantrums or deliberate misbehavior.



The body changes first

Stress activates the sympathetic nervous system, which shifts heart rate, skin conductance and skin temperature, often before anything is visible.



A warning creates time

Even a few minutes is enough for a parent or teacher to lower noise, offer a break, or move to a quieter space.

72%

of parents of autistic children said they would be willing to continuously monitor their child's physiological signals to better understand meltdowns

Koo et al., 2018

Sources: Mazefsky et al., 2013; McCorry, 2007; Goodwin et al., 2006; National Autistic Society

What earlier studies left open

Heart rate only

Khullar et al., 2019

Flagged abnormal heart rate from a Fitbit

One signal, one child

Low-cost wearable

Singh, 2022

Heart rate + skin response; 82% accuracy

One child, caregiver labels

Camera-based

Sindhoor Preetham et al., 2017

CNN recognized distress gestures (~92%)

Healthy adults, on camera only

Deep learning

Ali & Saravanan, 2023

CNN-LSTM reached 96% accuracy

Only deep models compared

The gap: few studies compare classical and deep learning models on the same multi-sensor data from several people, tested on people the model has never seen.

Two questions, three hypotheses

Research questions

1. Can wearable sensor data predict an autistic meltdown before it starts?
2. Which kind of model predicts it best?

H1 Deep learning will beat classical models, because it is better at patterns that change over time.

H2 The strongest predictors will be stress-response signals: heart rate, skin conductance and inter-beat interval.

H3 Using several sensors and several people will beat earlier single-signal or single-person work.

The data: eight people, one wristband, 43 meltdowns

8
autistic participants

~1 day
of recording each

690,854
one-second readings

43
labeled meltdown episodes

Empatica E4 signals: movement, blood volume pulse, skin conductance (EDA), heart rate, inter-beat interval, skin temperature

How each example was built



Windows slide forward every 30 seconds. The label only looks ahead, so no future data leaks into the input.

1.9%
of 22,536 windows (430) came before a meltdown. Always guessing “no” would be 98.1% accurate.

Dataset: Autism Together Hackathon Dataset (2022), released publicly on Kaggle; labels provided with the dataset; no demographic information included

Five models, tested on people they never saw

Feature-based models

Logistic Regression · Random Forest

- ~30 hand-built features per window: variability, range, heart-rate variability (RMSSD), skin-response peaks
- SMOTE oversampling on training data only

Deep learning models

CNN · LSTM · CNN-LSTM

- Read the raw 60-second signals and learn their own features
- Class imbalance handled by weighting the loss

Leave-one-subject-out validation (one of 8 rounds shown)

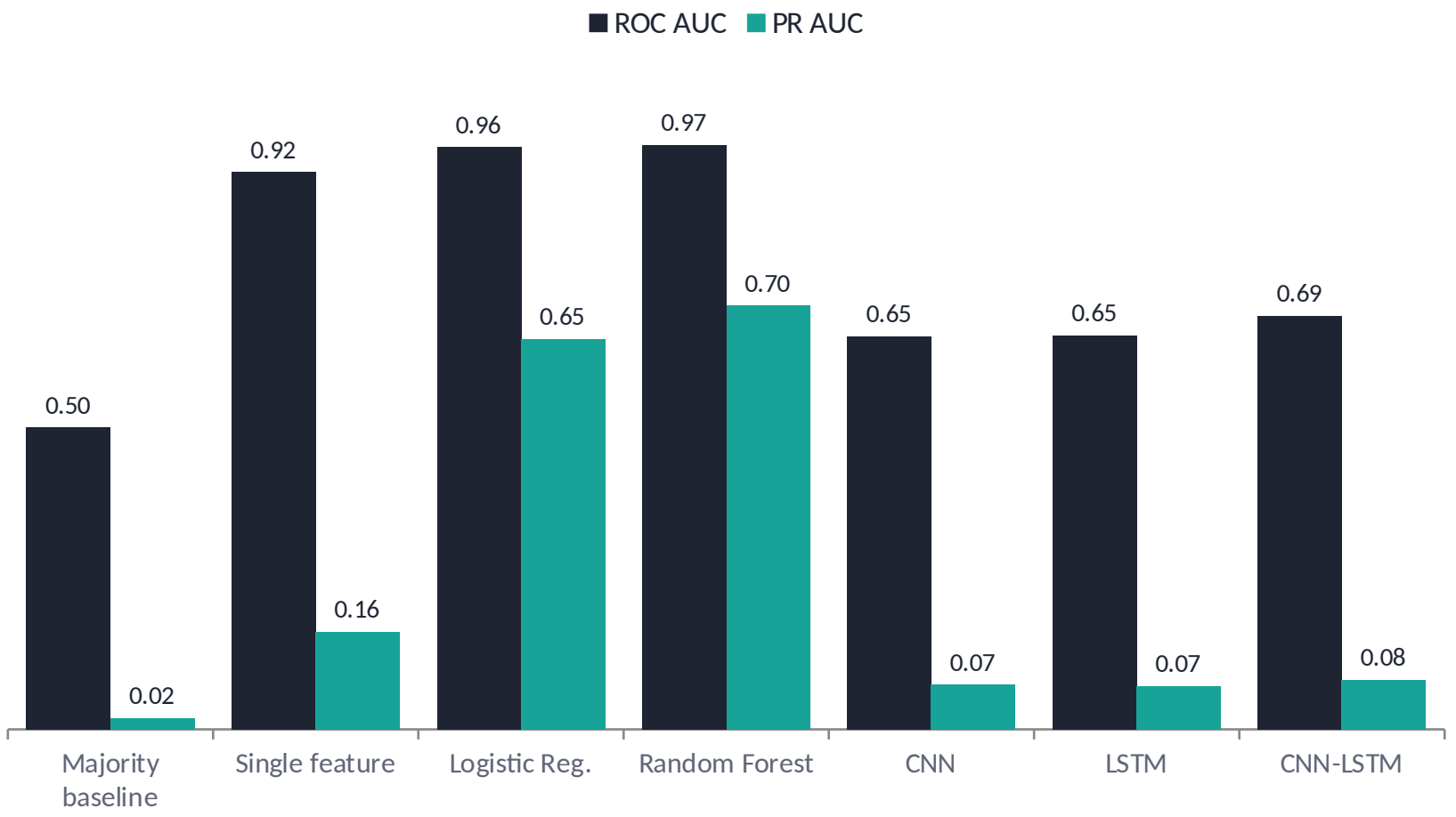


Baselines for comparison: always guess "no meltdown," and a single best feature with no model.

Why not a random split?

Overlapping windows would land in both training and test sets, and scores would look far better than they are.

Simple features beat deep learning



Random Forest

0.968

ROC AUC · PR AUC 0.702, about 37× the 0.019 a random guess gets

Best deep learning

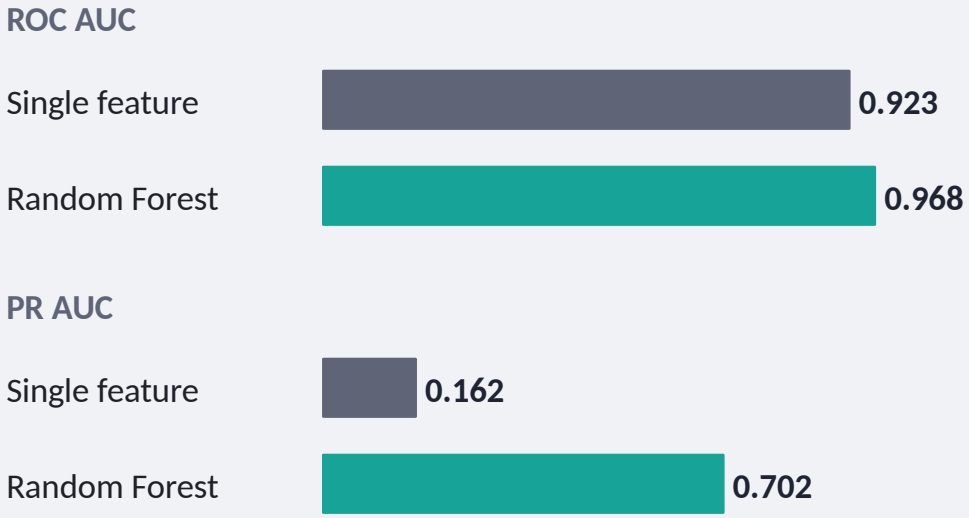
0.685

ROC AUC (CNN-LSTM) · PR AUC 0.082, about 4× chance

Mean across 8 held-out participants (leave-one-subject-out). ROC AUC: 0.5 = chance. PR AUC: chance = share of pre-meltdown windows (0.019).

Look past the headline number

One number, no model, already scores 0.923



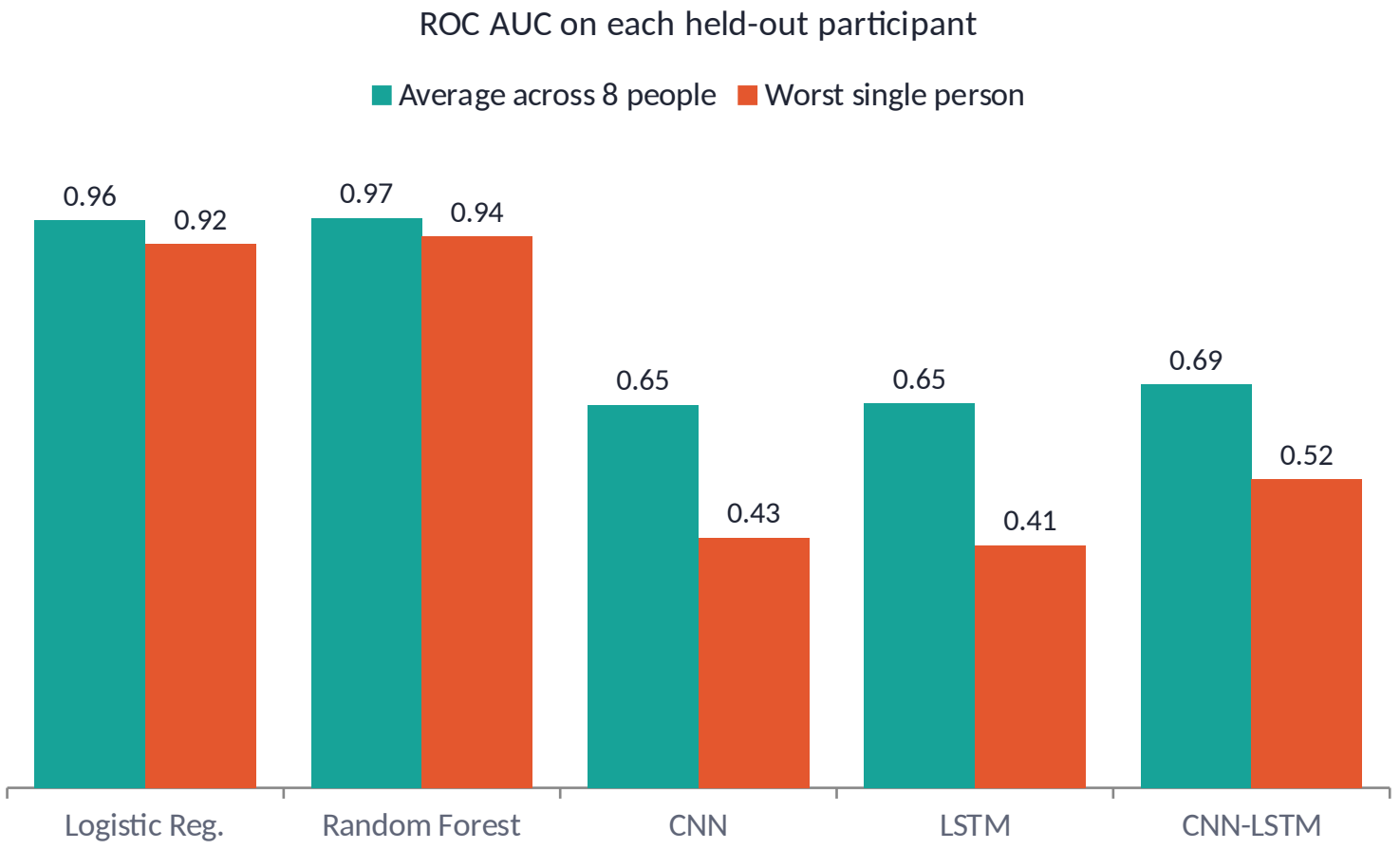
On ROC AUC they look close. On PR AUC, Random Forest is more than 4× better.

The best score may not be the best tool

	Precision	Recall
Random Forest Most alarms are real, but misses about half	0.692	0.547
Logistic Regression Catches far more, with more false alarms	0.446	0.826

Missing a meltdown costs more than a false alarm, so Logistic Regression may be the better choice in practice.

It has to work for the next person, too



0.937

Random Forest's lowest score on any participant

0.41–0.86

Range of deep learning scores across participants

0.916 → 0.717

CNN-LSTM score on the same fold when only the random starting point changed

The warning sign is unsteadiness, not a spike

Most important features, from Random Forest permutation importance

1 

Heart rate variability

Variance, standard deviation and range: the strongest by a clear margin

2 

Skin temperature variability

How much temperature moved within the minute

3 

Skin conductance (EDA) variability

Range, variance and skin-response peak height

– 

Movement (accelerometer)

Near chance: visible movement did not change before episodes

H2 supported: stress-response signals lead. The surprise is that variability matters more than level.



How far ahead?

1–5 min

Random Forest ROC AUC 0.983–0.990, no meaningful difference

10 min

Drops to 0.871: the signal fades

Single subject-wise split, classical models only

Why deep learning lost, and a trap I almost fell into

Too little data

About 330 pre-meltdown windows per training fold, each with 480 raw numbers. The network memorized instead of generalizing.

The answer is a variance

Random Forest is handed “heart rate variance” directly. A neural network has to compute it from 60 raw readings.

Examples aren’t independent

Ten overlapping windows lead up to each meltdown, so the models really saw 43 episodes, not 430 examples.

SMOTE and loss weighting were both tried; neither closed the gap.

The time-of-day trap

An early version scored much higher. Skin temperature drifts slowly over a day, so its average worked like a clock, and meltdowns don’t happen evenly across the day.

The model was learning the time, not the body. Removing average levels and keeping only variability fixed it.

No code error, no leakage, normal-looking metrics.

Limitations

Small sample

8 people and 43 episodes. Windows from one episode aren’t independent, so true uncertainty is wider than reported

Night-time episodes

13 of 43 episodes start between midnight and 7 a.m., during sleep, which is clinically unusual

Labels from the dataset

Meltdown labels came with the data; none were re-checked by trained observers

No demographics

Age, sex and support needs aren’t available, so I can’t tell who the results apply to

One prediction target

Main results use a 5-minute window; the horizon check used a single split

Offline only

Models were tested on stored recordings, not live on a wristband

What this study shows

H1	Deep learning wins	Rejected
H2	Stress signals lead	Supported
H3	Multi-sensor beats single	Consistent

- A warning signal exists in the 5 minutes before a meltdown, and simple features find it reliably.
- Start with hand-built features, not deep learning, at this sample size.
- Always report a trivial baseline, and check whether slow-drifting signals are acting as a clock.

Next step: more participants, with meltdown labels from trained observers.

Questions?

Dhruv Aggarwal

Selected references

Ali & Saravanan (2023). Upset or collapse detection system (CNN-LSTM).

Chawla et al. (2002). SMOTE. J Artif Intell Res, 16, 321–357.

Goodwin et al. (2006). Focus Autism Other Dev Disabil, 21(2).

Hochreiter & Schmidhuber (1997). Neural Computation, 9(8).

Khullar et al. (2019). Meltdown/tantrum detection system for ASD.

Koo et al. (2018). Archives of Design Research, 31(1), 37–55.

Mazefsky et al. (2013). J Am Acad Child Adolesc Psychiatry, 52(7).

McCorry (2007). Am J Pharm Educ, 71(4), 78.

Picard et al. (2016). Emotion Review, 8(1), 62–75.

Schmidt et al. (2018). WESAD. Proc. ICMI, 400–408.

Sindhoor Preetham et al. (2017). CNN recognition of distress gestures.

Singh (2022). Low-cost wearable for predicting autistic meltdowns.

Vos et al. (2023). J Biomedical Informatics, 148, 104556.

Dataset: Autism Together Hackathon Dataset (2022). Autistic Meltdown Prediction [Dataset]. Kaggle. https://www.kaggle.com/datasets/autismresearchcenter/autistic-meltdown-prediction