

applications of **deep learning**
in medical genetics

simon liu

medical genomics unit

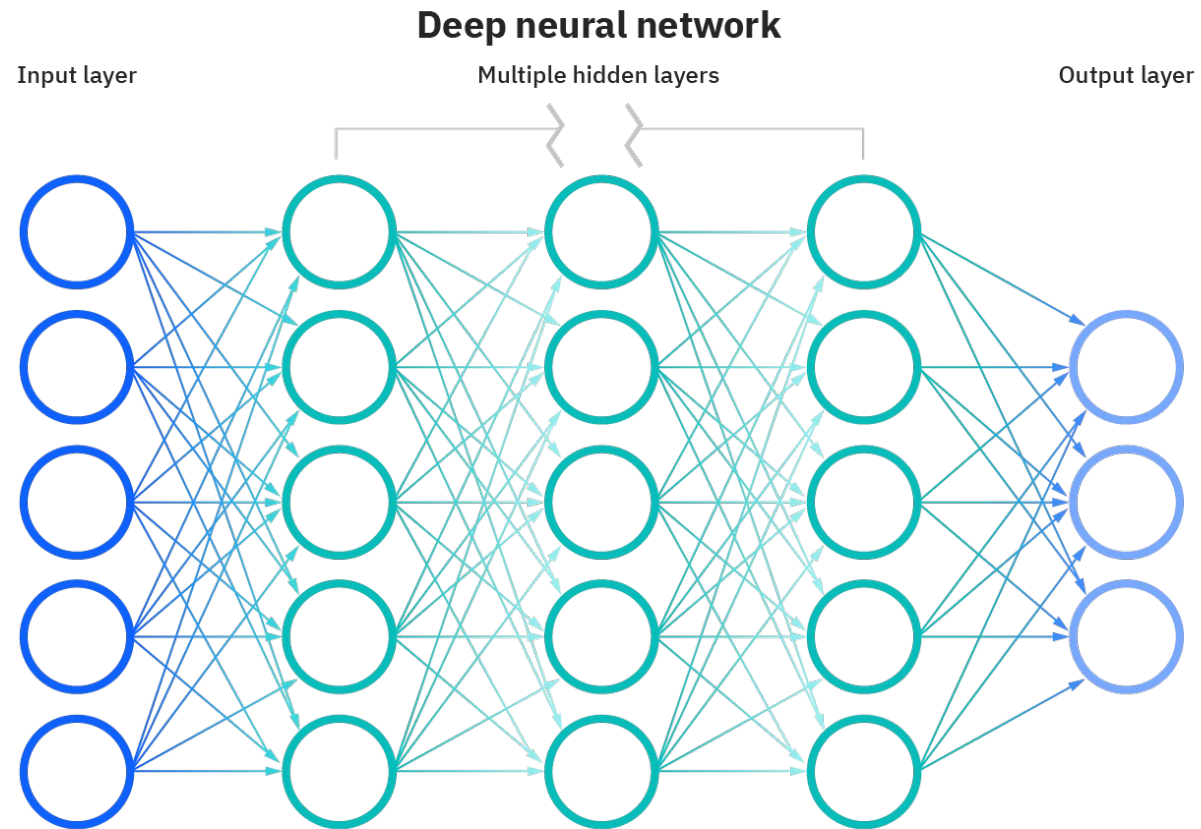
june 12, 2023



National Human Genome
Research Institute

—
The **Forefront**
of **Genomics**[®]
—

what is deep learning?



<https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>

automatic journal
abstract classification

journal article classification

- 14,002 article PMIDs collected initially
 - 306 were excluded due to unavailability (295), corrections (6), duplicates (5)
- random forest (RF) and Bidirectional Encoder Representations from Transformers (BERT)
- Snorkel2 used to aggregate classifier and labeling function predictions



journal article classification

Snorkel Model	TPR	FPR	FNR	TNR
LFs only	0.143	0.378	0.571	0.469
RFs only	0.214	0.000	0.786	1.000
BERTs only	1.000	0.055	0.000	0.945
LFs + RFs	0.143	0.378	0.643	0.469
LFs + BERTs	0.143	0.378	0.857	0.622
RFs + BERTs	1.000	0.055	0.000	0.945
LFs + RFs + BERTs	0.143	0.378	0.857	0.622

undiagnosed diseases program

undiagnosed diseases program

- ~950 patients admitted into the NIH UDP
 - 818 with HPO terms
 - 226 with genetic diagnoses

undiagnosed diseases program

	correct	incorrect	emp. acc.
karyotype	534	242	0.688
cma	501	275	0.646
sgd_test	440	336	0.567
gene_panel	485	291	0.625
mito_seq	511	265	0.659
exome	504	272	0.649
genome	538	238	0.693
genome_nih	406	370	0.523
biochem	456	320	0.588
male	403	373	0.519
pediatric	488	288	0.629



undiagnosed diseases program

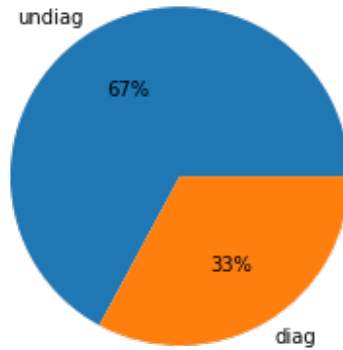
no prior genome

prior genome

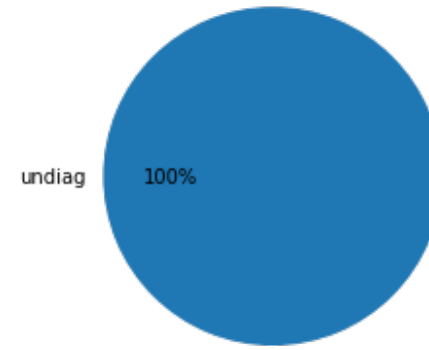
no nih genome

nih genome

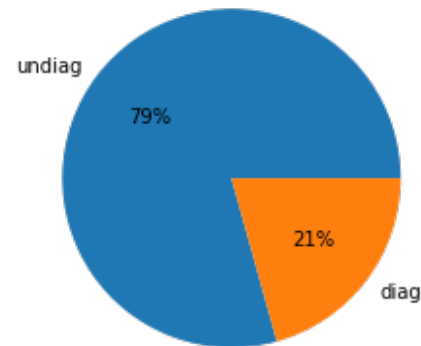
Patients w/o any genome



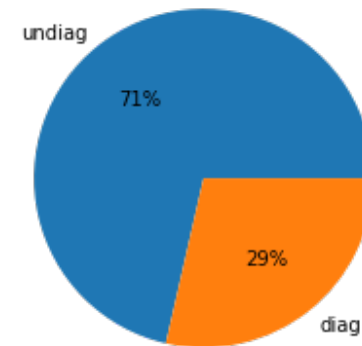
Patients with unanalyzed prior genomes



Patients w/o prior genome but w/ NIH genome

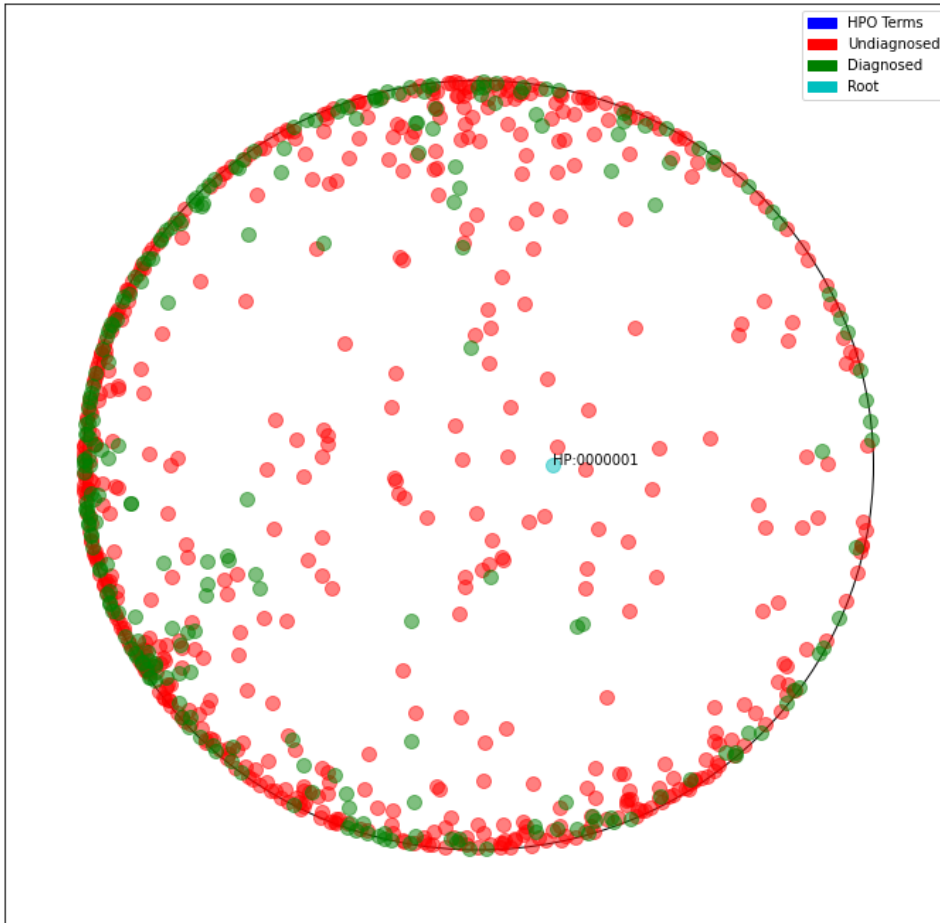


Patients with reanalyzed prior genomes

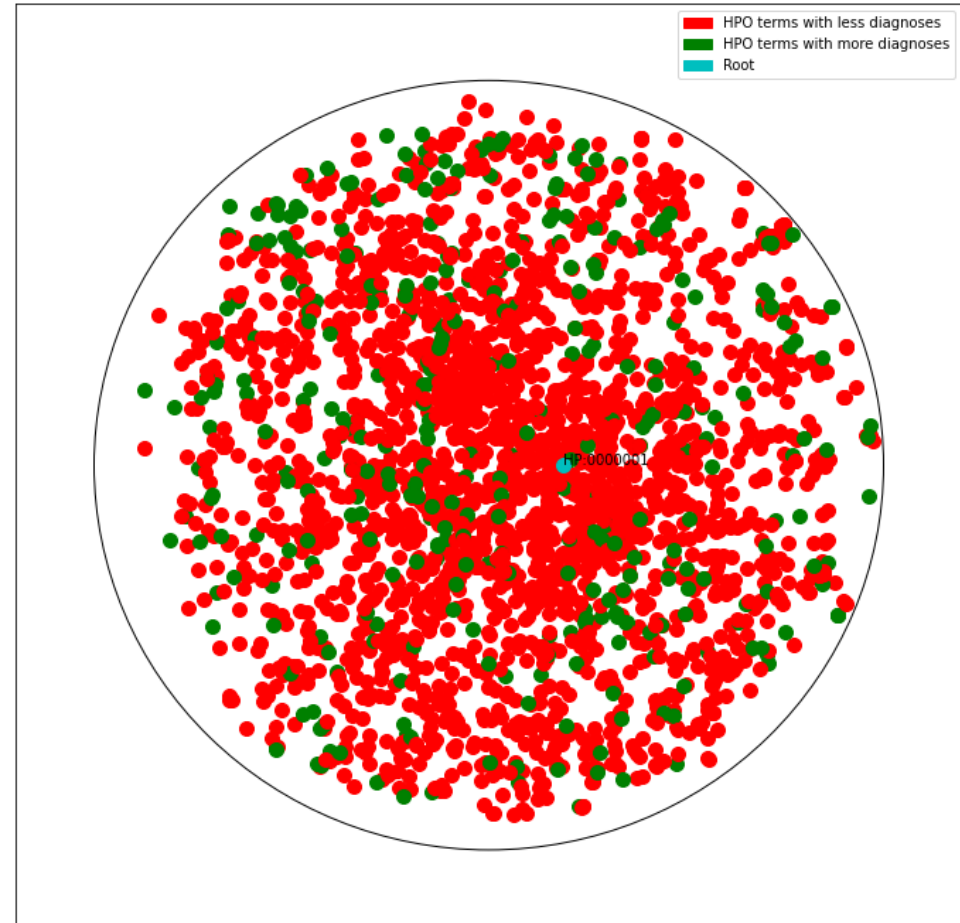


undiagnosed diseases program

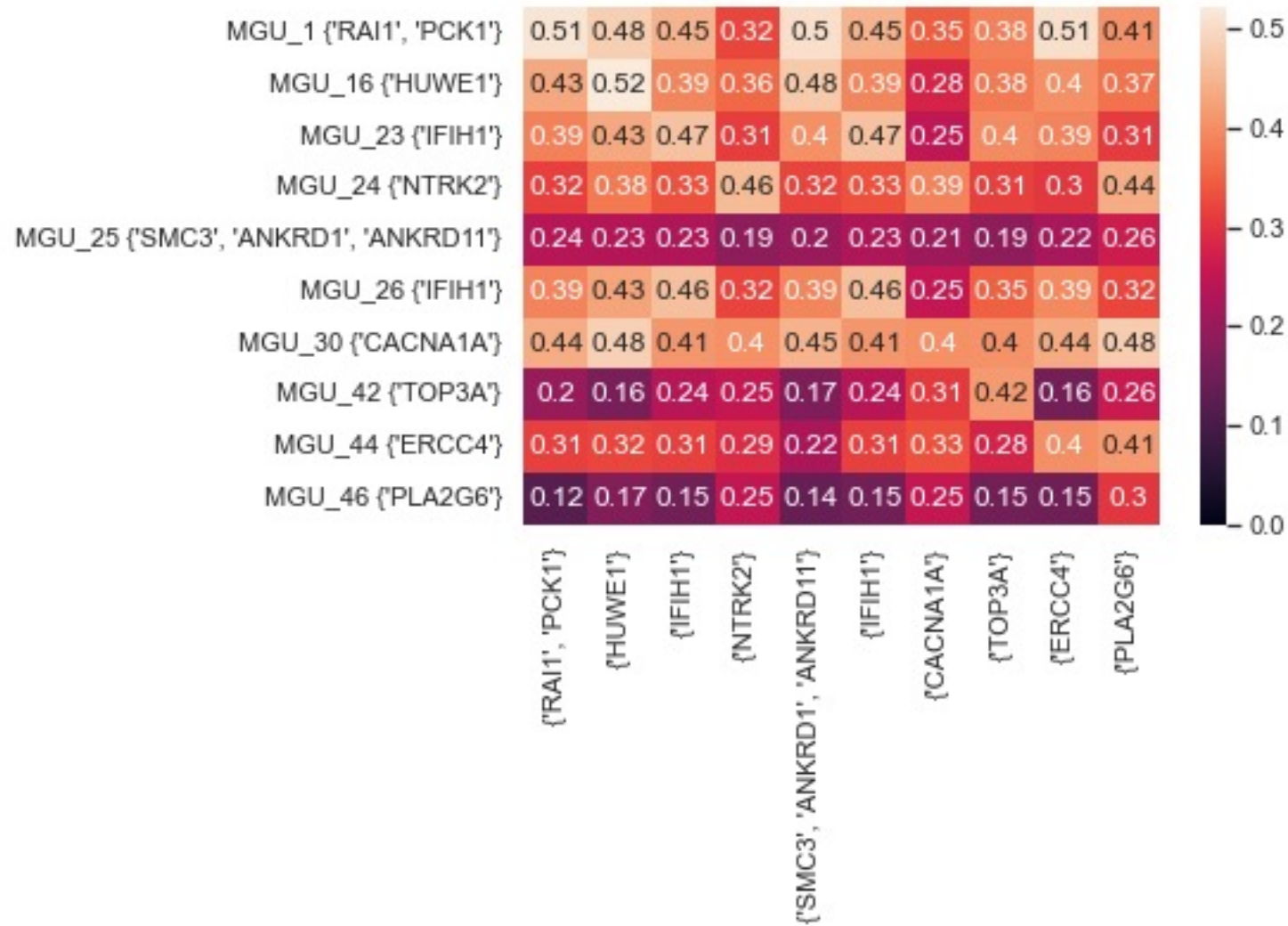
Poincare Ball Embedding of UDP Patients



Poincare Ball Embedding of UDP Patient HPO Terms

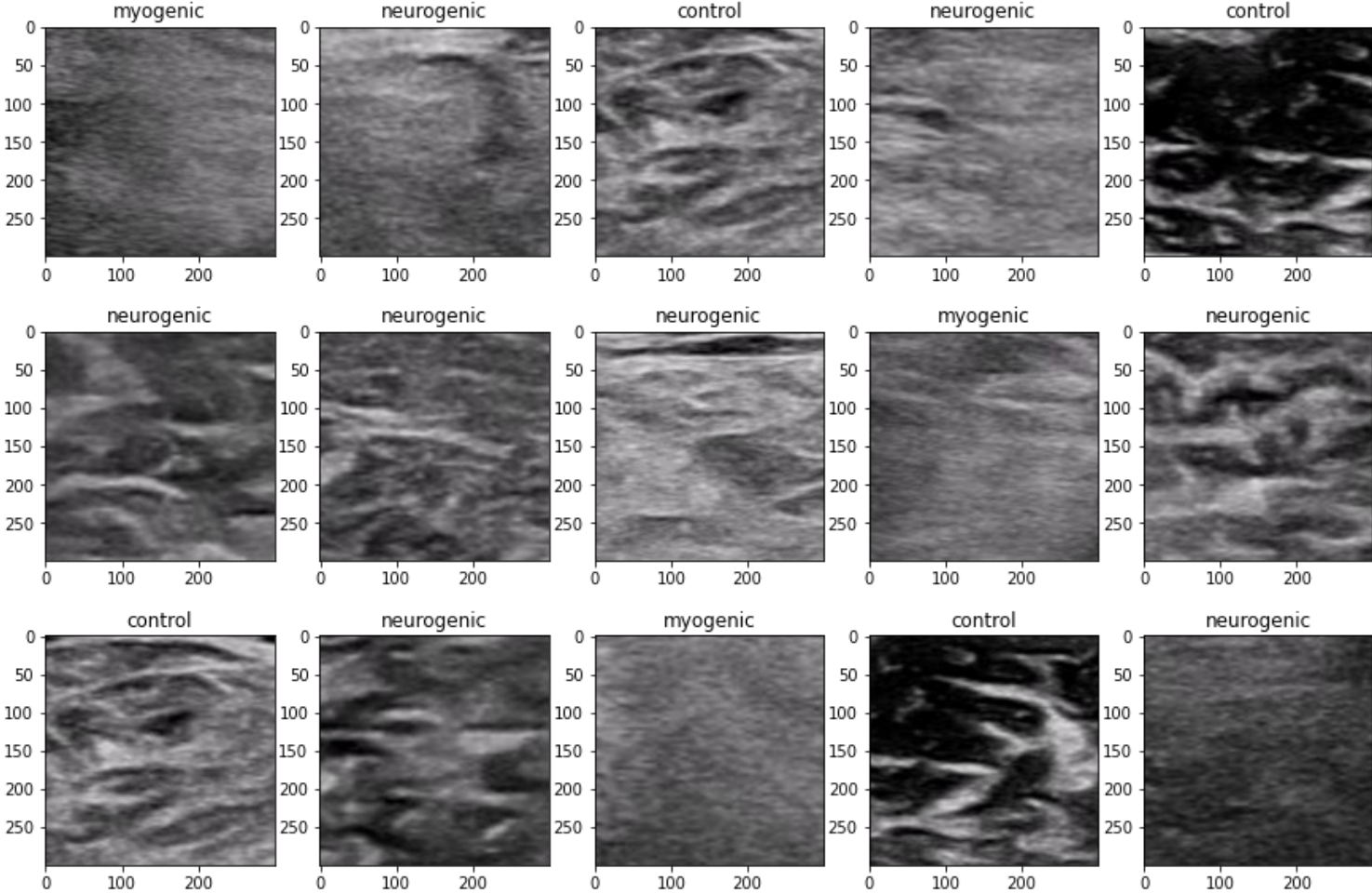


undiagnosed diseases program



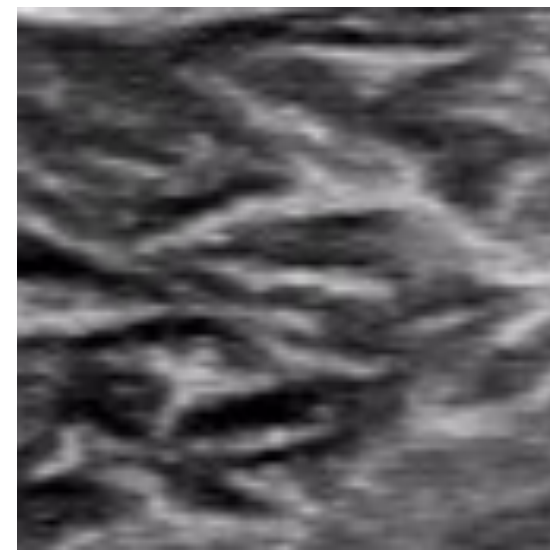
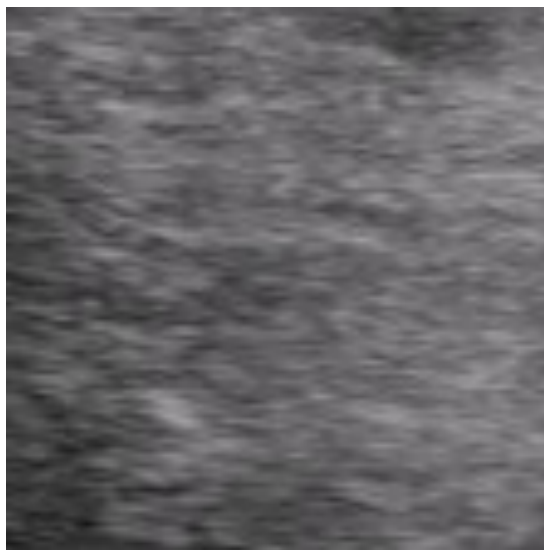
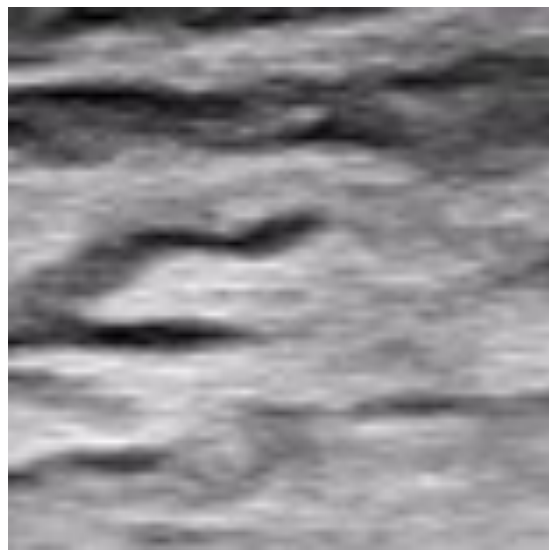
muscle ultrasound

muscle ultrasound

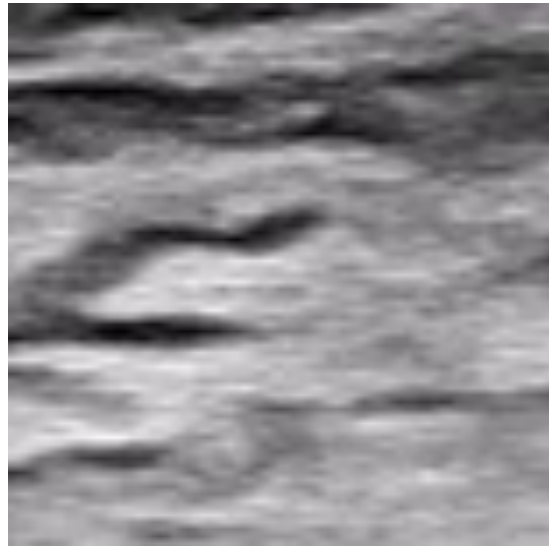


control	100
myogenic	50
neurogenic	102

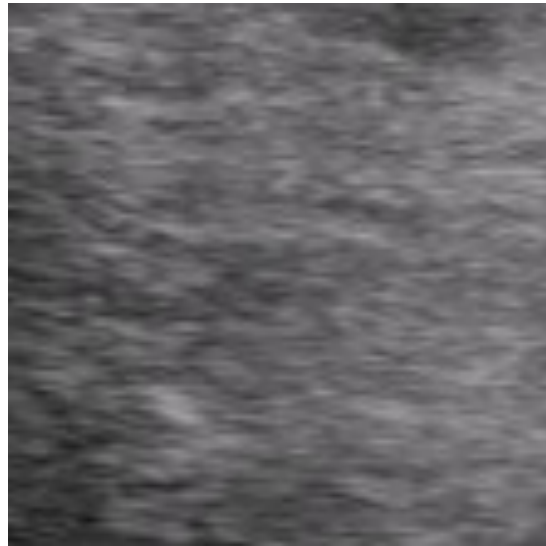
muscle ultrasound



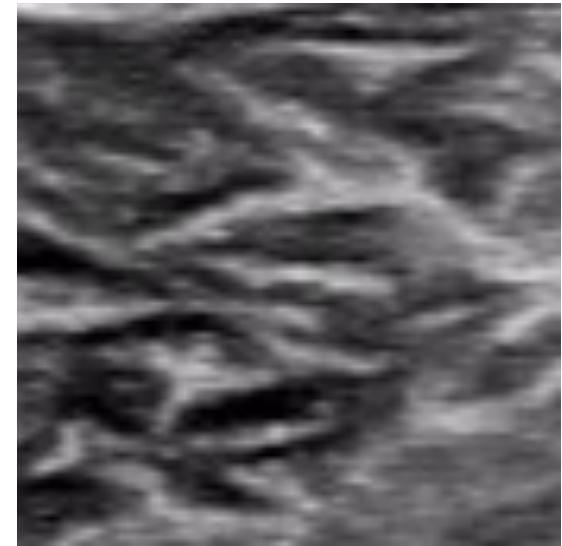
muscle ultrasound



neurogenic

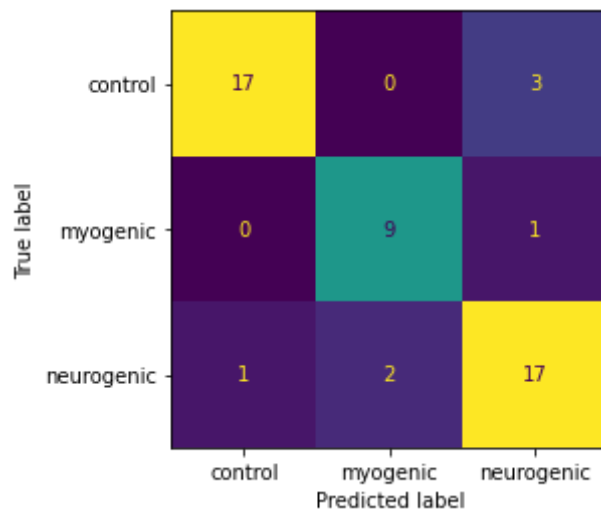
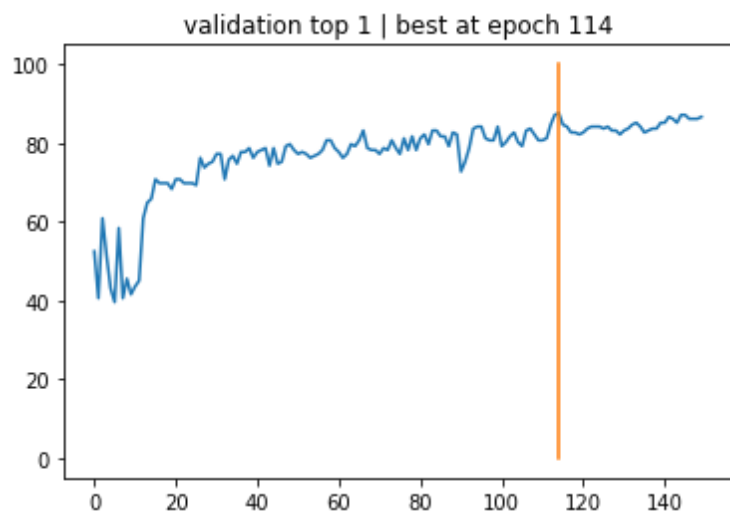
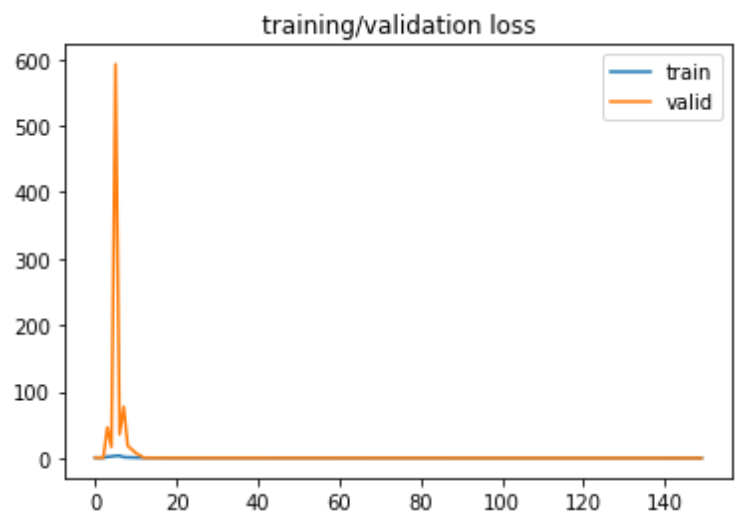


myogenic



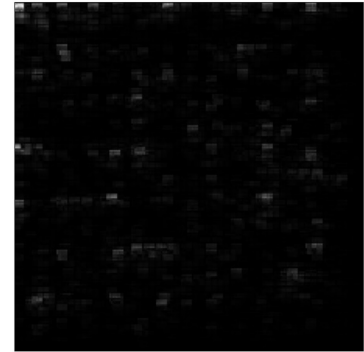
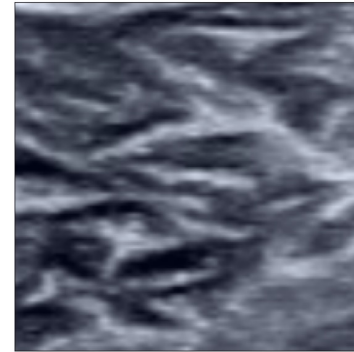
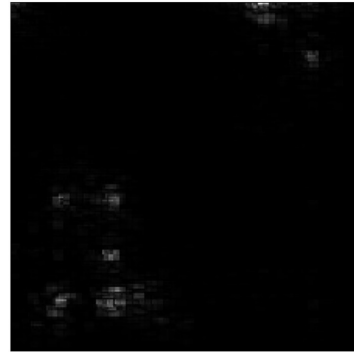
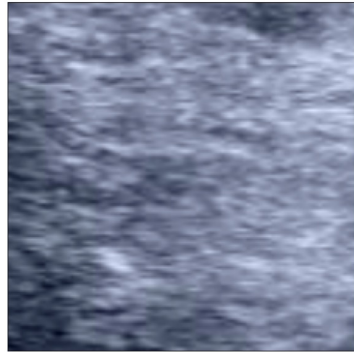
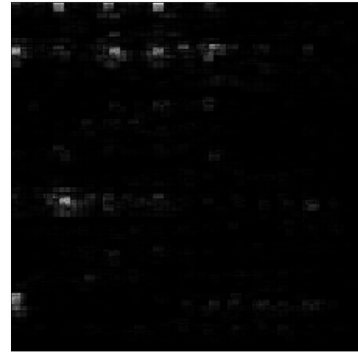
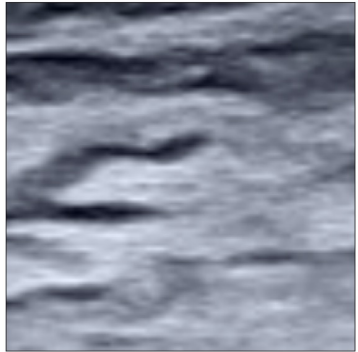
control

muscle ultrasound



0.8762 @ epoch 114
model: seresnet34

muscle ultrasound



neurogenic

myogenic

control



The **Forefront**
of **Genomics**[®]