Classification of Human Leukemias

Gordon Chan, Jonathan Stern, Shamanth Chedde, Quinton Zerafa

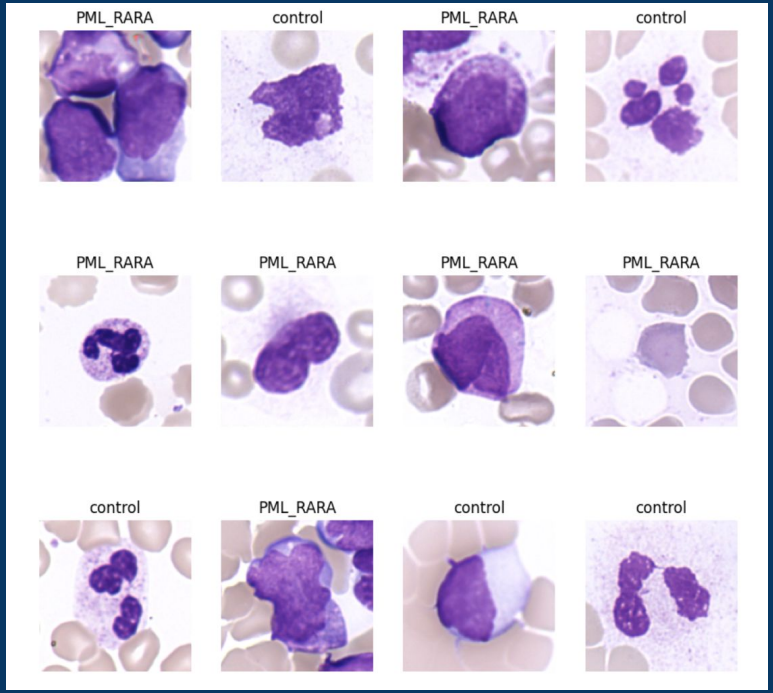
Cancer Image Archive

Blood Cell Images from Normal and Leukemia patients

4 Classes of Leukemia:

1. **Normal** (20,305 images)
2. **Leukemia with PML::RARA fusion** (11,584 images)
3. **Leukemia with NPM1 mutation** (17,710 images)
4. **Leukemia with CBFB::MYH11 fusion** (17,212 images)
5. **Leukemia with RUNX1::RUNX1T1 fusion** (14,403 images)

Normal (control) vs PML::RARA Leukemia



Sample Images

Normal (control) vs PML::RARA Leukemia

Input (64)

Conv2D (128)
Batch Normalization (128)
Max Pooling (128)

Conv2D (192)
Batch Normalization (192)
Max Pooling (192)

Flatten

Dense layer (128)

Dense layer (1)

Output

Learning Rate: 0.00001

Normal (control) vs Leukemia with RUNX1 fusion

Input (64)

Conv2D (256)
Batch Normalization (256)
Max Pooling (256)

Conv2D (192)
Batch Normalization (192)
Max Pooling (192)

Flatten

Dense layer (256)

Dense layer (1)

Output

Learning Rate: 0.00001

Normal (control) vs Leukemia with CBFB-MYH11 fusion

Input (32)

Conv2D (128)
Batch Normalization (128)
Max Pooling (128)

Conv2D (256)
Batch Normalization (256)
Max Pooling (256)

Conv2D (256)
Batch Normalization (256)
Max Pooling (256)

Flatten

Dense layer (192)

Dense layer (1)

Output

Learning Rate: 0.0001

Normal (control) vs Leukemia with NPM1 Mutation

Input (64)

Conv2D (256)
Batch Normalization (256)
Max Pooling (256)

Conv2D (192)
Batch Normalization (192)
Max Pooling (192)

Conv2D (192)
Batch Normalization (192)
Max Pooling (192)

Flatten

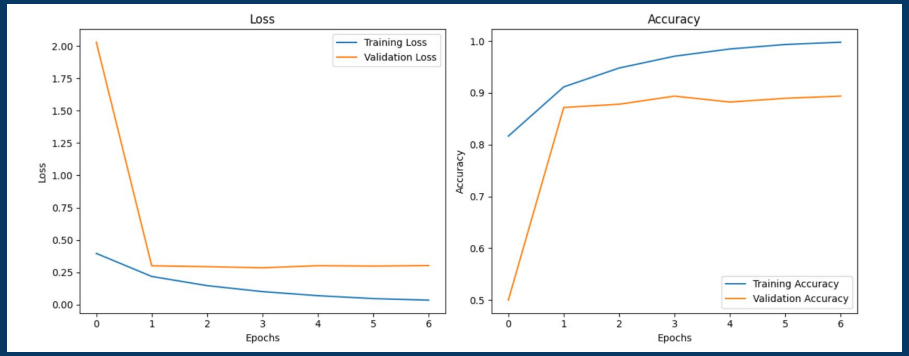
Dense layer (256)

Dense layer (1)

Output

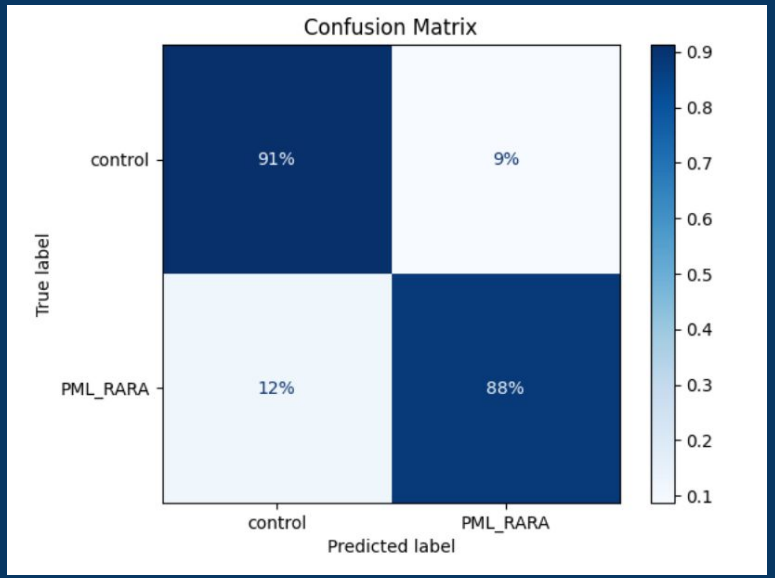
Learning Rate: 0.0001

Normal (control) vs PML::RARA Leukemia



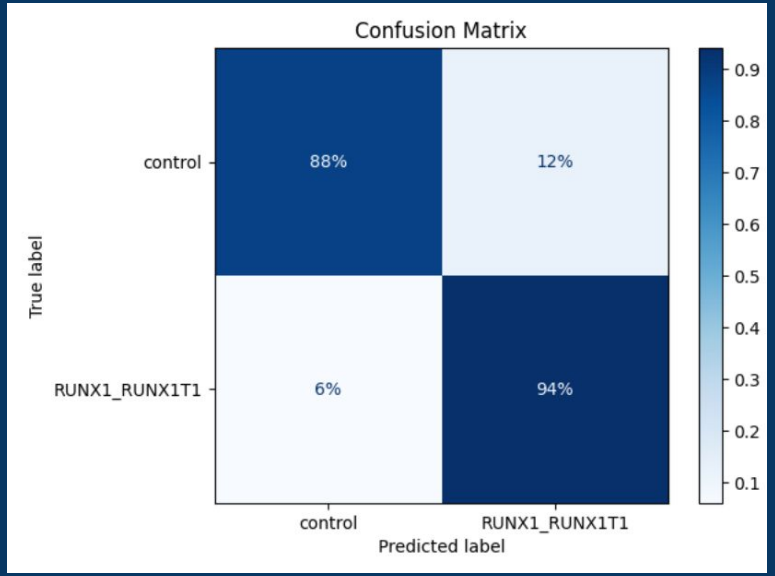
Training Loss (Post hyperparameter tuning)

Predictions	Precision	Recall	f1-Score	Support
Control	0.88	0.91	0.9	800
PML_RARA	0.91	0.88	0.9	800
Accuracy			0.9	1600
Macro Avg	0.9	0.9	0.9	1600
Weighted Avg	0.9	0.9	0.9	1600



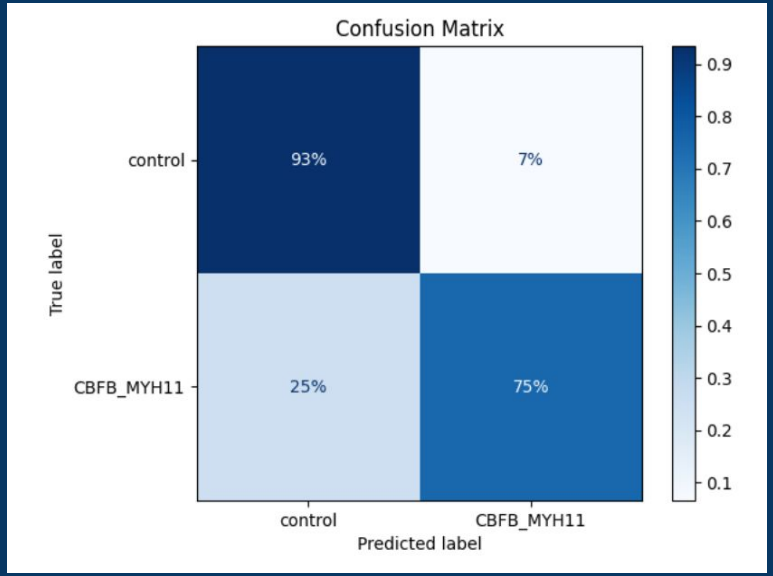
Normal (control) vs Leukemia with RUNX1 fusion

Predictions	Precision	Recall	f1-Score	Support
Control	0.94	0.88	0.91	800
RUNX1_RUNX1T1	0.89	0.94	0.91	800
Accuracy			0.91	1600
Macro Avg	0.91	0.91	0.91	1600
Weighted Avg	0.91	0.91	0.91	1600



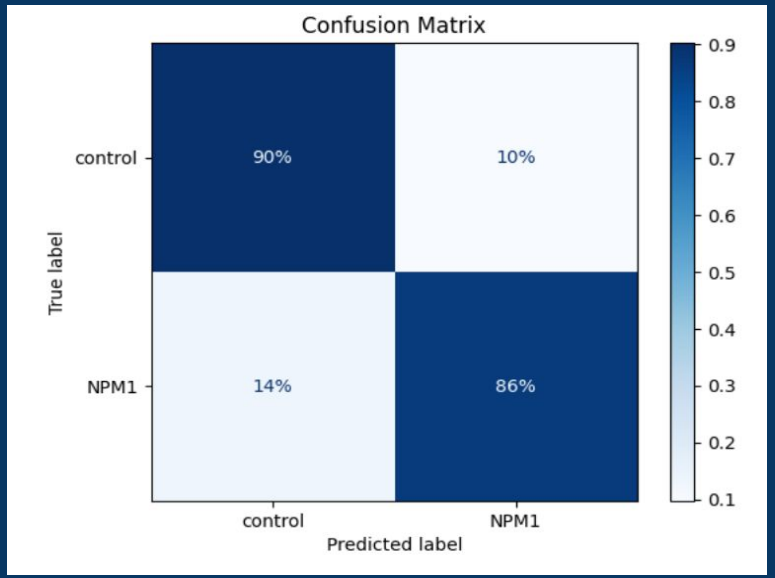
Normal (control) vs Leukemia with CBFB-MYH11 fusion

Predictions	Precision	Recall	f1-Score	Support
Control	0.79	0.93	0.86	800
CBFB-MYH11	0.92	0.75	0.83	800
Accuracy			0.84	1600
Macro Avg	0.85	0.84	0.84	1600
Weighted Avg	0.85	0.84	0.84	1600



Normal (control) vs Leukemia with NPM1 Mutation

Predictions	Precision	Recall	f1-Score	Support
Control	0.87	0.90	0.89	800
NPM1	0.90	0.86	0.88	800
Accuracy			0.88	1600
Macro Avg	0.88	0.88	0.88	1600
Weighted Avg	0.88	0.88	0.88	1600



Conclusion

Established 4 Image Classification models for for 4 different types of Human Leukemias.

F1 score of each model ~0.8 - 0.9.

Conclusion

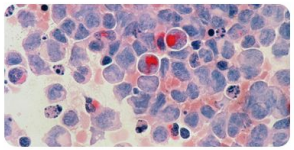
Each model was trained with ~5000 images from normal individual and from Leukemia patients.

Dataset and Notebooks are NOW available on Kaggle

https://www.kaggle.com/datasets/gchan357/human-aml-cytomorphology-dataset

Human Leukemia Cytomorphology Dataset

Dataset of Blood Cell Images from normal patients and patients with Leukemia



Data Card | Code (5) | Discussion (0) | Suggestions (0)

About Dataset

This dataset comprises blood cell images from four prevalent Acute Myeloid Leukemia (AML) subtypes with defining genetic abnormalities and typical morphological features according to the WHO 2022 classification: (i) APL with PML::RARA fusion, (ii) AML with NPM1 mutation, (iii) AML with CBFB::MYH11 fusion (without NPM1 mutation), and (iv) AML with RUNX1::RUNX1T1 fusion, as well as a control group of healthy stem cell donors.

Each folder under the root directory represent one classs. These are:

(1) normal patients (control)
(2) APL with PML::RARA fusion (PML_RARA)
(3) AML with NPM1 mutation (NPM1)
(4) AML with CBFB::MYH11 fusion without NPM1 mutation (CBFB_MYH11)
(5) AML with RUNX1::RUNX1T1 fusion (RUNX1_RUNX1T1).

Each sub-folder under the folder of each class represent images from a single patient with that Leukemia. Each patient's initial is the name of that folder.

Usability
7.50

License
Other (specified in description)

Expected update frequency
Never

Tags

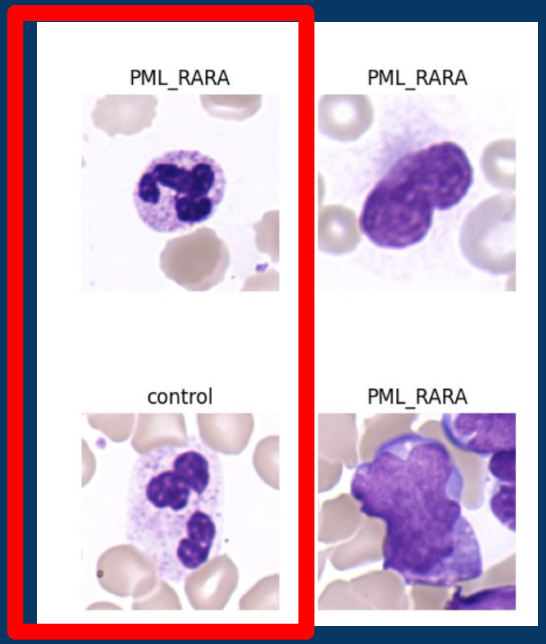
Cancer

Image Classification

Diseases

Leukemia patients sample may contain **normal** white blood cells.

- May in part explain **misclassification**?
- Datasets needs to be **curated**?



	World Health Organisation (WHO)[236]	Incidence	FAB type[237,238]	Incidence
1	AML with recurrent genetic abnormalities	30-40%		
	t(8;21)(q22;q22)		M2	25-30%
	inv(16)(p13q22) or t(16;16)(p13;q22)		M4eo	5-15%
	t(15;17)(q22;q12), (PML/RARα) and variants		M3	5-10%
	11q23 (MLL) abnormalities			
2	AML with multilineage dysplasia	10-15%		
	Following MDS or MDS/MPD			
	Without antecedent MDS or MPD, with dysplasia in at least 50% of cells in 2 or more myeloid lineages			
3	AML and MDS, therapy related	5-10%		
	Alkylating agents/radiation-related			
	Topoisomerase 2 inhibitor-related			
	Others			
4	AML, not otherwise categorised	30-40%		
	Minimally differentiated		M1	10-15%
	Without maturation		M0	2-5%
	With maturation			
	Acute myelomonocytic leukemia		M4	5-15%
	Acute monoblastic/monocytic leukemia		M5	15-25%
	Acute erythroid leukemia		M6	-3%
	Acute megakaryblastic leukemia		M7	5-10%
	Acute basophilic leukemia			
	Acute panmyelosis with myelofibrosis			
	Myeloid sarcoma			

MDS indicates myelodysplastic syndrome; MPD myeloproliferative syndrome

Models presented here may be used as a base model to train classification of other Leukemias using Transfer Learning.