General

ID[1]		
Use case name	Chromosome Segmentation and Deep Classification	
Application domain	Healthcare	
Deployment model	Hybrid or other (please specify)	
Status	PoC	
Scope[2]		
Objective(s)	• Automating Karyotyping of the chromosomes in cell spread images. • Segmentation of chromosomes in the images using non expert crowd.	
Narrative	Short description (not more than 150 words)	Karyotyping of the chromosomes micro-photographed under metaphase is done by characterizing the individual chromosomes in cell spread images. Currently, considerable effort and time is spent to manually segment out chromosomes from cell images, and classifying the segmented chromosomes. We proposed a method to segment out and classify chromosomes for healthy patients using a combination of crowdsourcing, preprocessing and deep learning, wherein the non-expert crowd from external crowdsourcing platform is utilized to segment out the chromosomes, which are then classified using deep neural network. Results are encouraging and promise to significantly reduce the cognitive burden of segmenting and karyotyping chromosomes.
	Complete description	Metaphase chromosome analysis is one of the primary techniques utilized in cytogenetics. Observations of chromosomal segments or translocations during metaphase can indicate structural changes in the cell genome, and is often used for diagnostic purposes. Karyotyping of the chromosomes micro-photographed under metaphase is done by characterizing the individual chromosomes in cell spread images. Currently, considerable effort and time is spent to manually segment out chromosomes from cell images, and classifying the segmented chromosomes into one of the 24 types, or for diseased cells to one of the known translocated types. Segmenting out the chromosomes in such images can be especially laborious and is often done manually, if there are overlapping chromosomes in the image which are not easily separable by image processing techniques. Many techniques have been proposed to automate the segmentation and classification of chromosomes from spread images with reasonable accuracy, but given the criticality of the domain, a human in the loop is often still required. In this paper, we present a method to segment out and classify chromosomes for healthy patients using a combination of crowdsourcing, preprocessing and deep learning, wherein the non-expert crowd from CrowdFlower is utilized to segment out the chromosomes from the cell image, which are then straightened and fed into a (hierarchical) deep neural network for classification. Experiments are performed on

<table>
<tr><td></td><td></td><td colspan="3">400 real healthy patient images obtained from a hospital. Results are encouraging and promise to significantly reduce the cognitive burden of segmenting and karyotyping chromosomes.</td></tr>
<tr><td>Stakeholders[3]</td><td colspan="4"></td></tr>
<tr><td>Stakeholders' assets, values[4]</td><td colspan="4"></td></tr>
<tr><td>System's threats and vulnerabilities[5]</td><td colspan="4"></td></tr>
<tr><td rowspan="3">Key performance indicators (KPIs)</td><td>ID</td><td>Name</td><td>Description</td><td>Reference to mentioned use case objectives</td></tr>
<tr><td>1</td><td>Classifier Accuracy</td><td>Without straightening and pre-processing, the average classification accuracy obtained was 68.5%. However, with preprocessing, the classification accuracy improved to 86.7%. These results are very likely to improve with more annotated training data for classification.</td><td></td></tr>
<tr><td>2</td><td>Annotation Completeness</td><td>35.9 chromosomes segmented out after crowd annotation, for 50 images having 46 chromosomes</td><td></td></tr>
<tr><td rowspan="5">AI features</td><td>Task(s)</td><td colspan="3">Recognition</td></tr>
<tr><td>Method(s)[6]</td><td colspan="3">Crowdsourcing and Deep learning</td></tr>
<tr><td>Hardware[7]</td><td colspan="3"></td></tr>
<tr><td>Topology[8]</td><td colspan="3"></td></tr>
<tr><td>Terms and concepts used[9]</td><td colspan="3">Deep learning, crowd sourcing, non-expert crowd, segmentation, karyotyping</td></tr>
<tr><td>Standardization opportunities/ requirements</td><td colspan="4"></td></tr>
<tr><td>Challenges and issues</td><td colspan="4">Crowd's job satisfaction
Spamming in annotated data</td></tr>
<tr><td>Societal concerns</td><td>Description</td><td colspan="3"></td></tr>
</table>

	SDGs[10]	Good health and well-being for people

Data (optional)

Data characteristics	
Description	The dataset comprised of 400 stained images with varying degrees of overlap between chromosomes, out of which 200 were kept for testing and the remaining for training and validation
Source[11]	Partner hospital
Type[12]	Images
Volume (size)	400
Velocity (e.g. real time)[13]	
Variety (multiple datasets)[14]	
Variability (rate of change)[15]	
Quality[16]	

Process scenario (optional)

Scenario conditions					
No.	Scenario name	Scenario description	Triggering event	Pre-condition[17]	Post-condition[18]

Training (optional)

Scenario name	Training				
Step No.	Event[19]	Name of process/Activity[20]	Primary actor	Description of process/activity	Requirement

Specification of training data[21]	

Evaluation (optional)

Scenario name	Evaluation				
Step No.	Event[22]	Name of process/Activity[23]	Primary actor	Description of process/activity	Requirement

Input of evaluation[24]	
Output of evaluation[25]	

Execution (optional)

Scenario name	Execution				
Step No.	Event[26]	Name of process/Activity[27]	Primary actor	Description of process/activity	Requirement

Input of Execution[28]	
Output of Execution[29]	

Retraining (optional)

Scenario name	Retraining				
Step No.	Event[30]	Name of process/Activity[31]	Primary actor	Description of process/activity	Requirement

Specification of retraining data[32]	

References

References						
No.	Type	Reference	Status	Impact on use case	Originator/organization	Link

1) Sharma, Monika & Saha, Oindrila & Sriraman, Anand & Hebbalaguppe, Ramya & Vig, Lovekesh & Karande, Shirish. (2017). Crowdsourcing for Chromosome Segmentation and Deep Classification. 786-793. 10.1109/CVPRW.2017.109.

Footnote

[1] Leave this cell blank.

[2] The scope defines the limits of the use case.

[3] Stakeholder involved in the scenario - examples are: type of organization; customers, 3rd parties; end users; humans; environment; negative stakeholders (attackers, criminals, etc).

[4] Assets and values that are valuable to the stakeholders and at the risk of being compromised by the AI system deployment – examples can include competitiveness; reputation or trust; fairness; safety; privacy; stability; etc.

[5] Threats and vulnerabilities can compromise the assets and values above. Examples are: different sources of bias; incorrect AI system use; new security threats; challenges to accountability; new privacy threats (hidden patterns).

[6] AI method(s)/framework(s) used.

[7] Hardware system used.

[8] Topology is the study of geometric forms differentiated by intersection and bifurcation. The term is used for the graphic aspects network architectures.

[9] Terms and concepts listed here can be used to extend the work of WG 1 (AWI 22989 and AWI 23053) as necessary.

[10] The Sustainable Development Goals (SDGs), otherwise known as the Global Goals, are a collection of 17 global goals set by the United Nations General Assembly. SDGs are a universal call to action to end poverty, protect the planet and ensure that all people enjoy peace and prosperity.
See URL for more details: http://www.undp.org/content/undp/en/home/sustainable-development-goals.html

[11] Origin of data, which could be from instruments, IoT, web, surveys, commercial activity, or from simulations.

[12] Structured/unstructured Images, voices, text, gene sequences, and numerical. Composite: time-series, graph-structured

[13] The rate of flow at which the data is created, stored, analysed, or visualized.

[14] Data from a number of domains and a number of data types. The wider range of data formats, logical models, timescales, and semantics complicates the integration of the variety of data.

[15] Changes in data rate, format/structure, semantics, and/or quality.

[16] Completeness and accuracy of the data with respect to semantic content as well as syntactical of the data (such as presence of missing fields or incorrect values)

[17] Describe which condition(s) should have been met before this scenario happens.

[18] Describe which condition(s) should prevail after this scenario happens. The post-condition may also define "success" or "failure" conditions.

[19] The event that triggers the step. This might be completion of the previous event.

[20] Action verbs should be used when naming activity.

[21] Training data can be further specified.

[22] The event that triggers the step. This might be completion of the previous event.

[23] Action verbs should be used when naming activity.

[24] Specify input of evaluation.

[25] Specify output of evaluation.

[26] The event that triggers the step. This might be completion of the previous event.

[27] Action verbs should be used when naming activity.

[28] Specify input of evaluation.

[29] Specify output of evaluation.

[30] The event that triggers the step. This might be completion of the previous event.

[31] Action verbs should be used when naming activity.

[32] Retraining data can be further specified.

