

Automating Systematic Reviews

Background

- SEBI – Supporting Evidence Based Interventions, staffed out of the Royal Dick and funded by BMGF. Collating animal health data from Google Scholar and other online databases and providing data visualisations
- Using PhD students from Informatics a machine learning tool was developed that could automate queries, reducing a three month manual task toward an end goal of three days, for a systematic review of literature
- The tool could be taught to automate other forms of systematic review

Background

- As a successful test case the tool was adapted to a small part of a literature review for the Microfluidics department in Engineering
- Engineering have 200 PhD and 100 Postdoc researchers who are expected to run ~2 reviews every three years, each review taking ~150 hours. Automation could reduce this to ~15 hours, a total time saving of 40 person-years in researcher time per year for Engineering alone.
- This could revolutionise research globally, saving hundreds of days of researcher time in nearly every university subject

Current State

- Continuing to work with SEBI to provide a full workflow that supports annotation, explainability, accuracy and features such as geolocation
- Working with Engineering to expand the proof of concept into a visual-based annotation tool that can generate training data
- Law: training a classifier to recognise and annotate a public database of Canadian refugee applications
- Medicine: Looking at the historical record of Cochrane reviews to provide a large set of training data

How Does the Tool Work?

- The tool uses Named Entity Recognition to recognise and infer terms that will be relevant to a given search criteria
- A subject matter expert will proceed with a manual review until sufficient annotations have been made for a classifier to be trained
- The classifier will classify papers for a given search into relevant, not relevant, and “low confidence” edge cases
- Low confidence results will be returned to the SME for annotation, creating a continuous improvement feedback loop

Signal – Confidence and Explainability

- The tool builds a “score” for each abstract it looks at, comparing it with the query
- Scores will fall into one of three categories : Include, Exclude, Low Confidence
- Low Confidence are borderline cases where the tool suspects a paper may relevant, but needs feedback to confirm
- “Signal” refers to the feedback from the tool on its confidence – **it understands the concept of accuracy or “truth”**
- The tool is able to explain its scoring for every paper
- The tool’s training data is available to show how the scoring was determined
- **Feedback from marking of Low Confidence papers allows the tool become more accurate**

10 documents have low confidence scores

Please examine each document and check whether it should be included or excluded. Once you have examined the documents, click 'Retrain Classifier' to add them to the training corpus and generate a new classifier.

Label*

Please provide a label for the new classifier.

Retrain Classifier ▶

Items per page: 10 ▾

« 1 »

Title	Abstract	DOI	Confidence	Country	Include/Exclude
Prevalence and phylogeography of	The study determined the prevalence and genetic population structure relationships of	https://doi.org/10.1017/S0031182022000713	-2.4459827691370153		<input type="checkbox"/> Excluded
Population structure, molecular characterization, and phylogenetic analysis of <i>Fasciola gigantica</i> from two locations in Uganda.	<i>Fasciola gigantica</i> is a major pathogen that causes fasciolosis in Africa. A recent study in Uganda demonstrated that <i>Fasciola</i> flukes were present in 65.7% of slaughtered cattle. However, molecular identification of <i>Fasciola</i> species has not yet been performed in the country. In the present study, 292 <i>Fasciola</i> flukes were collected from Kampala and Gulu, Uganda. The samples were identified as <i>F. gigantica</i> using a multiplex polymerase chain reaction (PCR) assay for phosphoenolpyruvate carboxykinase (pepck) and a PCR-restriction fragment length polymorphism (RFLP) assay for DNA polymerase delta (pold). A significant genetic difference between <i>F. gigantica</i> obtained from cattle slaughtered at Kampala and Gulu was observed by analyzing the mitochondrial markers NADH dehydrogenase subunit 1 (nad1) and cytochrome C oxidase subunit 1 (cox1). <i>Fasciola</i> collected from Gulu had a more diversified population than that collected from Kampala, probably because of differences in livestock management systems. One of the possible reasons for this observation is that cattle slaughtered in Gulu were reared under an extensive communal grazing system, which is suitable for maintaining parasite diversity, whereas cattle slaughtered in Kampala mainly originated from fenced/closed farms, which limits parasite diversity. However, the cause of the difference between these two locations was not clearly defined by the results of this study. The <i>F. gigantica</i> population from Uganda was related to that obtained from Zambia. A star-like phylogeny was detected in a median-joining network analysis, which indicated rapid population expansion and suggested that the <i>F. gigantica</i> populations from both countries are maintained by domestic ruminants in eastern Africa. Interestingly, the <i>F. gigantica</i> population from Uganda was not related to those from Egypt and Nigeria. The results of the present study suggest that <i>F. gigantica</i> populations in African countries are indigenous to each country or region.	https://doi.org/10.1016/j.meegid.2022.105359	-2.20453421106738	Uganda	<input type="checkbox"/> Excluded
Extended-Spectrum β-Lactamase-Producing	Beef cattle, one of the food-producing animals, are linked to humans through a shared environment and the food chain as a major source of animal protein. Antimicrobial drugs are readily accessible for use in food animal production in Nigeria. Beef cattle and abattoir environments harbor pathogenic bacteria such as We conducted a cross-sectional study among abattoir workers, beef cattle, and abattoir environments in Abuja and Lagos. Stool, cecal, and environmental samples were collected from apparently healthy workers, slaughtered cattle, and abattoir environments from May to December 2020. Data were collected electronically using open data kit app installed on a mobile phone. Antimicrobial susceptibility patterns were determined using the Kirby-Bauer disk diffusion method against a panel of 16 antimicrobial agents. Phenotypic and genotypic characterizations of the isolates were conducted. Data were analyzed with descriptive statistics. From 21.7% (This is the first report of	https://doi.org/10.3389/cimb.2022.869314	-2.023897790864813	Nigeria	<input type="checkbox"/> Excluded
Detection of	Cystic echinococcosis (CE) is a zoonotic disease of great importance worldwide. This study was conducted to determine the prevalence and antigenic profile of	https://doi.org/10.1007/s12639-022-01508-z	-1.8212082669788907		<input type="checkbox"/> Excluded
Africa-wide meta-analysis on the prevalence and distribution of human cystic echinococcosis and canine	Echinococcosis is a neglected zoonosis of increasing public health concern worldwide. According to the World Health Organization, 19,300 lives and 871,000 disability-adjusted life-years are lost globally each year because of cystic echinococcosis. Annual costs associated with cystic echinococcosis were estimated at US\$ 3 billion because of treatment of cases and losses in the livestock industry. We performed the random-effects model of meta-analysis using 51-year (1970–2021) data available from AJOL, Google Scholar, PubMed, Science Direct, Scopus and Web of Science. We also applied the Joanna Briggs Institute critical appraisal instrument for studies reporting prevalence data, the Cochran's Q-test, Egger's regression test and the single study deletion technique to assess study quality within study bias heterogeneity	https://doi.org/10.1186/s13071-022-05474-6	-0.5335094026222406	Sudan	<input type="checkbox"/> Excluded



Named Entities in SEBI

8	https://doi.org/10.1016/j.meegid.2022.105359	<p>Population structure, molecular characterization, and phylogenetic analysis of <i>Fasciola gigantica</i> from two locations in Uganda.</p>	<p>Patrick Vudriko, Richard Echodu, Michiyo Tashiro, Nozomi Oka, Kei Hayashi, Madoka Ichikawa-Seki</p>	<p>2022 Abstract</p> <p><i>Fasciola</i> [Entity: DISEASE, Score: 0.942] <i>gigantica</i> [Entity: DISEASE, Score: 1.000] is a major pathogen that causes fasciol [Entity: DISEASE, Score: 0.904] ##osis [Entity: DISEASE, Score: 0.993] in Africa. A recent study in Uganda [Entity: REGION, Score: 0.487] demonstrated that <i>Fasciola</i> [Entity: DISEASE, Score: 0.973] flukes [Entity: DISEASE, Score: 0.993] were present in 65.7% of slaughtered cattle [Entity: SPECIES, Score: 1.000] ##. However, molecular identification of Fasci [Entity: DISEASE, Score: 0.909] ##ola [Entity: DISEASE, Score: 0.797] species has not yet been performed in the country. In the present study, 29 [Entity: SAMPLE SIZE, Score: 0.996] ##2 <i>Fasciola</i> [Entity: DISEASE, Score: 0.969] flukes [Entity: DISEASE, Score: 0.994] were collected from Kampal [Entity: REGION, Score: 0.880] ##a [Entity: REGION, Score: 0.983] and Gulu [Entity: REGION, Score: 0.778] ##, Uganda [Entity: REGION, Score: 0.884] ##. The samples were identified as <i>F.</i> [Entity: DISEASE, Score: 0.943] <i>gigantica</i> [Entity: DISEASE, Score: 1.000] using a multiplex polymerase chain reaction (PCR [Entity: DIAGNOSTIC TEST, Score: 0.963] ##) [Entity: DIAGNOSTIC TEST, Score: 0.997] assay for phosphoenolpyruvate carboxykinase (pepck) and a PCR-restriction fragment length polymorphism (RF [Entity: DIAGNOSTIC TEST, Score: 0.958] ##LP [Entity: DIAGNOSTIC TEST, Score: 0.940] ##) assay for DNA polymerase delta [Entity: DIAGNOSTIC TEST, Score: 0.718] (pold). A significant genetic difference between <i>F.</i> [Entity: DISEASE, Score: 0.999] <i>gigantica</i> [Entity: DISEASE, Score: 1.000] obtained from cattle [Entity: SPECIES, Score: 1.000] slaughtered at Kampal [Entity: REGION, Score: 0.926] ##a [Entity: REGION, Score: 0.692] and Gulu [Entity: REGION, Score: 0.916] was observed by analyzing the mitochondrial markers NADH dehydrogenase subunit 1 (<i>nad1</i>) and cytochrome C oxidase subunit 1 (<i>cox1</i>). Fasci [Entity: DISEASE, Score: 0.942] ##ola collected from Gulu [Entity: REGION, Score: 0.804] had a more diversified population than that collected from Kampal [Entity: REGION, Score: 0.850] ##a [Entity: REGION, Score: 0.597] ##, probably because of differences in livestock [Entity: SPECIES, Score: 1.000] management systems. One of the possible reasons for this observation is that</p>	-0.414487155158403	<p>Uganda: 5, Zambia: 1, Egypt: 1, Nigeria: 1, United States of America: 1, Australia: 1, Indonesia: 1</p>	Uganda
---	---	---	--	--	--------------------	--	--------

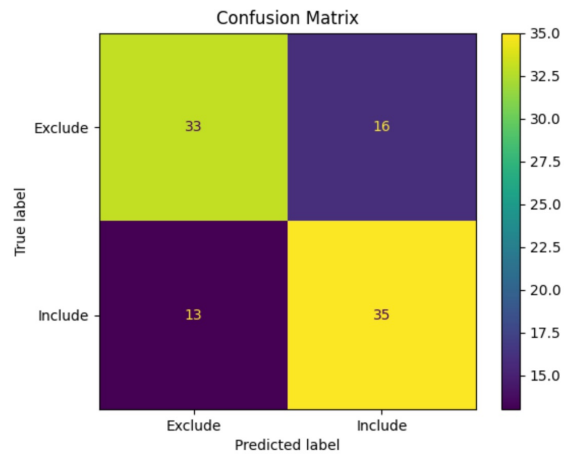
Status

Classifier training is complete

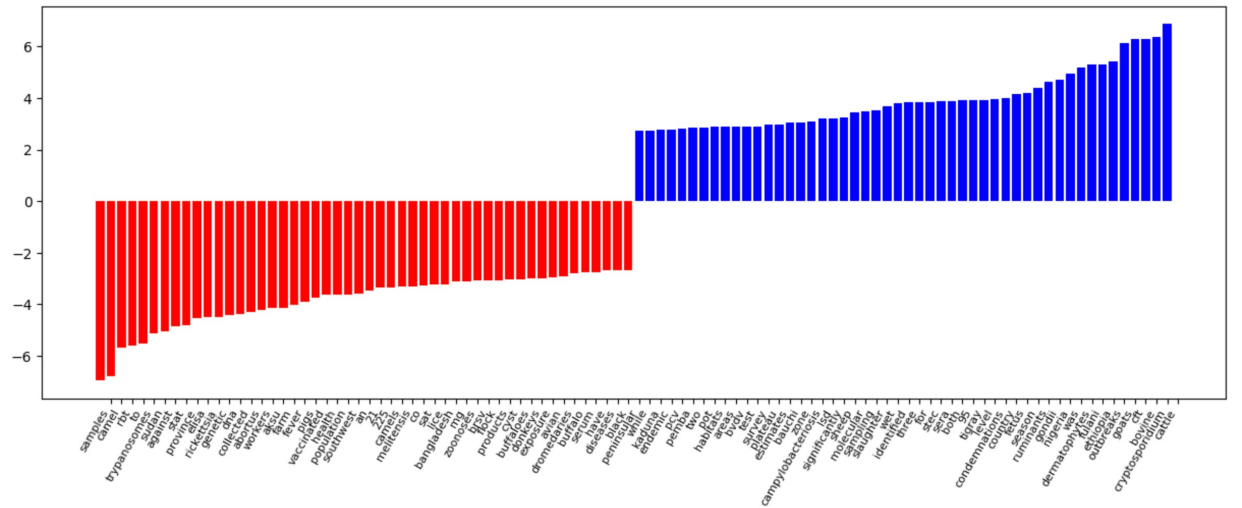
Classification Report [Download](#)

	Precision	Recall	F1-Score	Support
Exclude	0.72	0.67	0.69	49
Include	0.69	0.73	0.71	48
Accuracy			0.7	97
Macro avg	0.7	0.7	0.7	97
Weighted avg	0.7	0.7	0.7	97

Confusion Matrix [Download](#)



Coefficients [Download](#)



Geolocation

atus Settings Low Confidence Documents Locations

Title
Africa-wide meta-analysis on the prevalence and distribution of human cystic echinococcosis and canine Echinococcus granulosus infections.

Authors
Solomon Ngutor Karshima, Musa Isyaku Ahmed, Nuhu Bala Adams, Abdullahi Alhaji Magaji, Musa Zakariah, Konto Mohammed

Country Mentions
Sudan: 1, Libya: 1, United States of America: 1, Indonesia: 1

DOI
10.1186/s13071-022-05474-6

Abstract
Echinococc [Entity: DISEASE, Score: 0.999] Fosis [Entity: DISEASE, Score: 0.490] is a neglected zoonosis of increasing public health concern worldwide. According to the World Health Organization, 19,300 lives and 871,000 disability-adjusted life-years are lost globally each year because of cystic echinococc [Entity: DISEASE, Score: 0.978]. Fosis [Entity: DISEASE, Score: 0.734] Annual costs associated with cystic echinococc [Entity: DISEASE, Score: 0.940] Fosis [Entity: DISEASE, Score: 0.916] were estimated at US\$ 3 billion because of treatment of cases and losses in the livestock [Entity: SPECIES, Score: 0.997] industry. We performed the random-effects [Entity: STATISTICAL ANALYSIS, Score: 0.988]

Microfluidics

$n_i(\mathbf{r}, t_+)$ and $n_i(\mathbf{r} + \mathbf{c}_i, t_+)$, corresponding to velocities \mathbf{c}_i and \mathbf{c}_i ($\mathbf{c}_i = -\mathbf{c}_i$) parallel to the link connecting \mathbf{r} and $\mathbf{r} + \mathbf{c}_i$ the notation $n_i(\mathbf{r}, t_+) = n_i(\mathbf{r}, t) + \Delta_i(\mathbf{r}, t)$ is used to indicate the post-collision distribution, (2.7). The velocity of the boundary node \mathbf{u}_b is determined by the solid-particle velocity \mathbf{U} , angular velocity $\boldsymbol{\Omega}$, and centre of mass \mathbf{R} ,

$$\mathbf{u}_b = \mathbf{U} + \boldsymbol{\Omega} \times (\mathbf{r} + \frac{1}{2}\mathbf{c}_i - \mathbf{R}). \quad (2.8)$$

By exchanging population density between n_i and $n_{i'}$ the local momentum density can be modified to match the velocity of the solid-particle surface at the boundary node, without affecting either the mass density or the stress, which depend only on the sum $n_i + n_{i'}$. The precise form for the boundary-node collision operator is

$$\left. \begin{aligned} n_i(\mathbf{r} + \mathbf{c}_i, t + 1) &= n_i(\mathbf{r} + \mathbf{c}_i, t_+) + 2a_i^{c_i} \rho \mathbf{u}_b \cdot \mathbf{c}_i, \\ n_{i'}(\mathbf{r}, t + 1) &= n_{i'}(\mathbf{r}, t_+) - 2a_{i'}^{c_{i'}} \rho \mathbf{u}_b \cdot \mathbf{c}_i \end{aligned} \right\} \quad (2.9)$$

As a result of the boundary-node interactions (2.9), forces are exerted on the solid particles,


$$\mathbf{f}(\mathbf{r} + \frac{1}{2}\mathbf{c}_i, t + \frac{1}{2}) = 2[n_i(\mathbf{r}, t_+) - n_{i'}(\mathbf{r} + \mathbf{c}_i, t_+) - 2a_i^{c_i} \rho \mathbf{u}_b \cdot \mathbf{c}_i] \mathbf{c}_i; \quad (2.10)$$

thus momentum is exchanged locally between the fluid and the solid particle, but the combined momentum of solid and fluid is conserved. The forces and torques on the solid particle, obtained by summing $\mathbf{f}(\mathbf{r} + \frac{1}{2}\mathbf{c}_i)$ and $(\mathbf{r} + \frac{1}{2}\mathbf{c}_i) \times \mathbf{f}(\mathbf{r} + \frac{1}{2}\mathbf{c}_i)$ over all the boundary nodes associated with a particular particle, are then used to update the particle velocity and angular velocity, according to the laws of Newtonian mechanics. The mass and moment of inertia of the particle are preassigned, depending on the

Dataset

First page

Keywords

Immigration and Refugee Board of Canada Refugee Protection Division			Commission de l'immigration et du statut de réfugié du Canada Section de la protection des réfugiés	
RPD File No. / N° de dossier de la SPR : MB2-06059				
Private Proceeding / Huis clos				
Reasons and decision – Motifs et décision				
Claimant(s)	XXXX XXXX XXXX		Demandeur(e)(s) d'asile	
Date(s) of hearing	October 28, 2013	Date(s) de l'audience		
Place of hearing	Montréal, Quebec	Lieu de l'audience		
Date of decision and reasons	December 16, 2013	Date de la décision et des motifs		
Panel	Stéphane Morin	Tribunal		
Counsel for the claimant(s)	M ^e Marie-José Blain	Conseil(s) du (de la/des) demandeur(e)(s) d'asile		
Designated representative	N/A	Représentant(e) désigné(e)		
Counsel for the Minister	N/A	Conseil du (de la) ministre		

2013 CanLII 92087 (CA IRB)

REFUGEE PROTECTION DIVISION / EMPLOYMENT / EXCLUSION / ART. 1F(A) / CRIMES AGAINST HUMANITY / COMPLICITY / PROFILE / FAILURE TO DISASSOCIATE / MALE / NEGATIVE / PARTICULAR SOCIAL GROUP / FAMILY / REFUGEE SUR PLACE / WELL-FOUNDED FEAR OF PERSECUTION / FEMALE / CHILDREN / POSITIVE / ETHIOPIA

Features

Date(s) of hearing

Place of hearing

Date of decision and reasons

Panel

Counsel for the claimant(s)

>>> Other text data:

- + country reports
- + international conventions
- + local regulation, guidelines and laws

Cochrane

- Gold standard for medical literature reviews
- Every two years a review must be updated with the latest literature – in other words the review must be repeated
- 4-5x time consuming as ordinary research paper

But:

- Data on historical reviews is meticulous, machine-readable and can be ingested automatically
- Repetition every two years is key motivator for automation
- Key determinant for acceptance is data quality – e.g. double blind trials – which are terms a classifier can select for

ED*i*NA