

Evaluasi instrumen Multiple Machine Learning yang diaplikasikan pada studi epidemiologi infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii* di masyarakat pulau kecil dan terluar Indonesia

Junaidi, Umi Cahyaningsih, Syibbran Mulaesyi

Pendahuluan

Entamoeba histolytica (*E. histolytica*) sebagai agent patogen utama penyebab penyakit amebiasis menjadi isu penting kesehatan masyarakat dunia karena berhubungan dengan tingkat kesakitan dan kematian yang tinggi. Kematian yang disebabkan oleh penyakit ini menempati peringkat ketiga tertinggi setelah malaria dan scistosomiasis.(1) (2) Di samping itu, parasit protozoa intestinal ini juga dikelompokkan dalam salah satu agen parasit patogen pertahanan hayati prioritas kategori B, yakni agen yang dapat menginfeksi banyak orang melalui air dan atau makanan serta berpotensi untuk dijadikan sebagai senjata biologis.(3) Vektor mekanik seperti lalat dan kontak langsung pada aktivitas seksual oro-anal atau ano-genital menjadi jalur penularan lain yang sudah banyak dilaporkan.(4) (5)

Parasit intestinal dari famili Amoebidae ini memiliki spesies identik morfologi dengannya yaitu *E. dispar* dan *E. moshkovskii*. Berbeda dengan *E. histolytica*, kedua jenis amoeba tersebut merupakan amoeba non-patogen yang secara mikroskopis sering ditemukan secara bersamaan. Walaupun kapasitas virulensi *E. dispar* dan *E. moshkovskii* sangat berbeda dengan *E. histolytica*.(6) (7) Namun, keberadaan perlu diwaspadai karena karakteristik habitat dan cara penularannya tidak berbeda dengan *E. histolytica*, hal ini menyebabkan faktor risiko penularannya juga sama. walaupun infeksi *E. histolytica* sebenarnya dalam studi seperti ini dapat ditelusuri lebih lanjut dengan berbagai metode identifikasi lain, penemuan faktor risiko infeksi parasit patogen dan non patogen yang identik ini perlu di eksplorasi dan di carikan instrumen yang valid untuk memprediksi kejadian infeksi agen parasit ini di masa yang akan datang (8).

Indonesia sebagai negara kepulauan memiliki banyak pulau besar dan kecil. Pulau-pulau yang berada di bagian terluar Indonesia sebagian besar adalah pulau-pulau kecil dan sebagian diantaranya adalah pulau berpenduduk. Salah satu pulau kecil Indonesia yang perainnya berbatasan langsung dengan perairan negara lain adalah Pulau Weh. Pulau yang terletak di ujung paling barat Indonesia ini berbatasan langsung dengan perairan India, Malaysia dan Thailand (9). Sebagai mana wilayah lain di Indonesia, masyarakat yang tinggal di Pulau Weh berhadapan dengan masalah kesehatan masyarakat yaitu masalah sebaran pemukiman penduduk yang tidak merata dan terkonsentrasi di daerah pesisir atau lokasi strategis lainnya. Kondisi ini secara tidak langsung berdampak pada kualitas sanitasinya sehingga memungkinkan penyebaran dan penularan berbagai agen penyakit berbasis lingkungan dan salah satunya adalah *E. histolytica*. (10) (11)

Machine learning merupakan salah satu sistem artificial intelligence applications yang mengeksplorasi analisis dan konstruksi algoritma memungkinkan sistem tersebut untuk belajar dari data dan membuat prediksi atau keputusan tanpa perlu diprogram secara eksplisit.(12) (13) Penelitian ini bertujuan untuk mengestimasi prevalensi, menganalisis faktor risiko, dan mengeksplorasi model-model dalam Multiple Machine Learning sehingga menemukan model yang ideal dalam memprediksi kejadian infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii* pada masyarakat pulau Weh dan Masyarakat Pulau-pulau kecil dan terluar Indonesia lainnya.

Dalam studi ini kami akan menggunakan empat model machine learning dan empat metode sampling data. Dataset tersebut dibagi menjadi 10% data uji dan 90% data training. Kami akan membandingkan performa model-model ini menggunakan metrik AUROC,

AUPRC, F1 score, akurasi, CV Mean (rata-rata cross-validation) dan CV Std (standar deviasi cross-validation) untuk menentukan kombinasi model dan metode sampling yang paling efektif. Selain itu, setiap model machine learning juga akan dilatih tanpa menggunakan metode sampling data untuk memahami performa dasar mereka. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan baru tentang cara terbaik untuk memprediksi infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii* yang pada akhirnya dapat membantu dalam pengembangan strategi pencegahan dan penanggulangan penyakit amebiasis.

Metode

Jenis dan Sampel Penelitian

Penelitian ini adalah studi epidemiologi yang diaplikasikan artificial intelligence applications untuk mengevaluasi berbagai model machine learning. Hasil evaluasi ini nantinya dapat digunakan untuk memprediksi kejadian infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii* pada masyarakat pulau weh dan wilayah lain nantinya. Sampel dalam penelitian ini adalah masyarakat pulau Weh yang berumur ≥ 10 tahun. Responden penelitian dipilih secara Non-probability sampling yang berjumlah 335 responden. Kriteria inklusi sampel adalah penduduk asli Kota Sabang, dapat berkomunikasi dengan baik, dan bersedia menjadi subjek penelitian. Kriteria eksklusi adalah sampel tidak menyelesaikan seluruh proses pengambilan data.

Pemeriksaan dan Pengukuran

Pemeriksaan sampel

Pengambilan sampel tinja dimulai dengan memberikan *Stool Container* pada masing-masing responden. Pada hari berikutnya, peneliti mengumpulkan pot yang sudah di isi sampel tinja dan diperiksa di laboratorium. Identifikasi kompleks *Entamoeba histolytica/dispar/moshkovskii* dilakukan secara mikroskopis. Identifikasi ini dimulai dengan skrining keberadaan amoeba usus menggunakan teknik pemeriksaan langsung dan konsentrasi dengan larutan formol eter atau zinc sulfate.(14) (15). Sampel yang positif ditemukan pada salah satu teknik skrining tersebut, kemudian dibuat preparat kering dan diwarnai dengan teknik pewarnaan wheatley's trichrome,(16) dan selanjutnya diidentifikasi karakteristik sel untuk membedakan kompleks *Entamoeba histolytica/dispar/moshkovskii* dengan amoeba usus lainnya menggunakan Mikroskop Olympus Binokuler CX 23 dan camera digital microscope Merk Dino-Lite Series AM4025X.

Pengukuran perilaku kesehatan dan sanitasi lingkungan

Penelusuran informasi tentang perilaku pencegahan dan penanggulangan penyakit yang disebabkan oleh *E. histolytica* dilakukan dengan teknik wawancara terstruktur. Sedangkan Penelusuran informasi tentang sumber dan kecukupan air bersih diperoleh dari jawaban responden atau anggota keluarga. Data penelitian yang diperoleh dari teknik observasi adalah kondisi rumah, jamban keluarga, Sarana pembuangan air limbah (SPAL), dan Sarana pembuangan sampah (SPS) menggunakan instrumen check list pengamatan rumah dan sanitasi yang diterbitkan oleh Departemen Kesehatan RI 1999 dan telah dimodifikasi menyesuaikan dengan penelitian ini.

Manajemen dan Analisis Data

Data diolah dan dianalisis menggunakan software statistik dan Machine Learning. Dataset Machine Learning dibagi menjadi 10% data uji dan 90% data pelatihan. Pembagian ini dilakukan untuk memastikan bahwa model Machine Learning ini dapat belajar dari sebagian

besar data (data pelatihan) dan kemudian diuji pada data yang belum pernah dilihat sebelumnya (data uji) untuk memastikan keakuratan dan generalisabilitas model. Pengujian ini dilakukan pada masing-masing empat model Machine Learning dan metode data sampling. Pada tahap selanjutnya, setiap model dilatih baik dengan dan tanpa metode sampling data. Hasil dari training kemudian dianalisis menggunakan metrik seperti AUROC, AUPRC, F1 score, akurasi, CV std, dan CV avg.

Hasil

Kejadian dan prevalensi infeksi kompleks *Entamoebahistolytica/dispar/moshkovskii*

Hasil identifikasi mikroskopis terhadap 335 sampel tinja responden, ditemukan kejadian infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii* sebanyak 89 kasus dengan prevalensi temuan kasus ini sebesar 26,6%.

Prilaku kesehatan, sanitasi dan korelasi dengan kejadian infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii*

Tabel 1. Prilaku kesehatan, sanitasi korelasinya dengan kejadian infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii* pada masyarakat pulau Weh Provinsi Aceh

Aspek sanitasi dasar	N = 335 n (%)	Terinfeksi dengan kompleks <i>Eh/Ed/Em</i> *		P value
		n Tidak (%)	n Ya (%)	
Prilaku kesehatan				
▪ Perilaku mencuci tangan menggunakan air dan sabun pada saat makan atau menyentuh makan				0,027*
Sering	92 (27,5)	60 (65,2)	32 (34,8)	
Jarang	243 (72,5)	186 (76,5)	57 (23,5)	
▪ Perilaku memerhatikan kebersihan makanan, penjual, dan warung sebelum membeli				0,322
Selalu memerhatikannya	150 (44,8)	114 (76,0)	36 (24,0)	
Kurang memerhatikannya	119 (35,5)	132 (71,4)	53 (28,6)	
▪ Perilaku mencuci bersih sayuran dan buah-buahan yang dimakan segar atau bersama kulitnya				0,157
Selalu melakukannya.	130 (38,8)	91 (70,0)	39 (30,0)	
Jarang melakukannya	205 (61,2)	155 (75,6)	50 (24,4)	
▪ Perilaku memotong kuku 1x seminggu				0,237
Sering	183 (54,6)	131 (71,6)	52 (28,4)	
Sesekali	152 (45,4)	115 (75,7)	37 (24,3)	
▪ Perilaku mencuci tangan dengan sabun setiap kali selesai BAB				0,352
Sering (selama ini selalu dilakukannya).	173 (51,6)	125 (72,3)	48 (27,7)	
Sesekali (selama ini dimelakukan bersamaan dengan waktu mandi).	162 (48,4)	121 (74,7)	41 (25,3)	
▪ Perilaku mengakses informasi kesehatan				0,472
Sering	123 (36,7)	88 (71,5)	35 (28,5)	
Jarang	146 (43,6)	112 (76,7)	34 (23,3)	
Sangat jarang	66 (19,7)	46 (69,7)	20 (30,3)	
▪ Perilaku mencari pengobatan				0,687
Ke praktek Dokter/Puskesmas/Klinik Swasta/Rumah Sakit.	112 (33,4)	79 (70,5)	33 (29,5)	
Apotik, Toko Obat/Praktek Mantri/Bidan	177 (52,8)	133 (75,1)	44 (24,9)	
Orang Pintar, Dukun, Ustaz	46 (13,7)	34 (73,9)	12 (26,1)	
Kondisi sanitasi dasar				
▪ Sumber air bersih				0,002*

PDAM	100 (29,9)	83 (33,9)	17 (19,1)	0,001*
Sumur	111 (33,1)	85 (34,6)	26 (29,2)	
Sumber lainnya	124 (37,0)	78 (31,7)	46 (51,7)	
▪ Kecukupan air bersih				
Cukup	60 (17,9)	33 (55,0)	27 (45,0)	0,667
Kurang mencukupi	275 (82,1)	213 (77,5)	62 (22,5)	
▪ Jamban keluarga (JK)				
Saniter	167 (49,9)	119 (71,3)	48 (28,7)	
Tidak saniter	106 (31,6)	80 (75,5)	26 (24,5)	0,064
Tidak memiliki jamban keluarga	62 (18,5)	47 (75,8)	15 (24,2)	
▪ Sarana pembuangan air limbah (SPAL)				
Ada, tertutup	94 (28,1)	75 (30,5)	19 (21,3)	
Tidak ada, terbuka	241 (71,9)	171 (69,5%)	70 (78,7)	0,361
▪ Sarana pembuangan sampah (SPS)				
Ada, kedap air, tertutup dan sampah dibuang ke penampungan besar dibakar/dikubur	55 (16,4)	42 (76,4)	13 (23,6)	
Tidak ada, sampah dibuang di tanah dan dibiarkan terbuka	280 (83,6)	204 (72,9)	76 (27,1)	

Tabel 1 terlihat bahwa, dari enam aspek perilaku pencegahan dan usaha mencari pengobatan yang di analisis dalam penelitian ini, hanya variabel pencegahan yaitu mencuci tangan dengan air dan sabun pada saat makan atau menyentuh makan yang korelasi dengan kejadian infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii*. Korelasi ini merupakan korelasi negatif yang secara deskriptif dapat dijelaskan bahwa dari 92 responden yang menyatakan sering mencuci tangan dengan air mengalir dan sabun pada saat makan atau menyentuh makan, ditemukan sebanyak 34,8% responden yang terinfeksi dengan kompleks *Entamoeba histolytica/dispar/moshkovskii*. Sedangkan dari 243 responden yang menyatakan jarang mencuci tangan dengan air mengalir dan sabun pada saat makan atau menyentuh makan sebanyak hanya 23,5% responden yang terinfeksi dengan kompleks *Entamoeba histolytica/dispar/moshkovskii*. Dua dari Lima aspek sanitasi yang di teliti dalam studi ini, yaitu sumber dan kecupan air bersih diketahui berkorelasi dengan kejadian infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii*. Studi aspek sanitasi ini memiliki kesamaan dengan analisis sebelumnya. Dimana, terlihat korelasi negatif pada variabel kecukupan air bersih. Dimana, dari 60 responden atau rumah tangga yang menyatakan cukup memperoleh air bersih ditemukan sebanyak 45.0% responden positif terinfeksi dengan kompleks *Entamoeba histolytica/dispar/moshkovskii*. Sedangkan dari 275 responden atau rumah tangga yang menyatakan kurang cukup memperoleh air bersih ditemukan sebanyak 22,5 % responden yang positif terinfeksi dengan kompleks *Entamoeba histolytica/dispar/moshkovskii*.

Berikut adalah hasil dari penelitian kami menggunakan berbagai model Machine Learning dan metode sampling data:

Table 1 metrik AUROC, AUPRC, Acuraccy, F1 Score, Cross Validation Mean and standard deviation yang diaplikasikan pada data survey epidemiologi kejadian infeksi kompleks *Entamoeba histolytica/dispar/moshkovskii* pada masyarakat pulau Weh Provinsi Aceh

Model	Sampler	AUROC	AUPRC	Accuracy	F1	CV Mean	CV Std
DecisionTreeClassifier	TomekLinks	0.811	0.687	0.765	0.636	0.563	0.100
DecisionTreeClassifier	No sampling	0.765	0.664	0.765	0.636	0.543	0.098
DecisionTreeClassifier	ADASYN	0.758	0.619	0.706	0.545	0.621	0.075
GradientBoostingClassifier	No sampling	0.731	0.695	0.647	0.000	0.571	0.059
GradientBoostingClassifier	SMOTE	0.720	0.597	0.647	0.538	0.749	0.135
GaussianNB	No sampling	0.682	0.616	0.647	0.455	0.560	0.059

DecisionTreeClassifier	BorderlineSMOTE	0.680	0.556	0.618	0.552	0.613	0.081
DecisionTreeClassifier	SMOTE	0.680	0.556	0.618	0.552	0.660	0.088
GradientBoostingClassifier	TomekLinks	0.663	0.494	0.647	0.000	0.644	0.056
GradientBoostingClassifier	ADASYN	0.663	0.440	0.618	0.552	0.715	0.134
GaussianNB	TomekLinks	0.652	0.496	0.647	0.538	0.601	0.059
GradientBoostingClassifier	BorderlineSMOTE	0.644	0.433	0.647	0.538	0.742	0.103
SVC	BorderlineSMOTE	0.644	0.527	0.618	0.435	0.768	0.105
SVC	SMOTE	0.625	0.516	0.647	0.500	0.772	0.100
SVC	ADASYN	0.621	0.470	0.588	0.364	0.740	0.094
GaussianNB	SMOTE	0.614	0.560	0.588	0.462	0.678	0.131
SVC	No sampling	0.610	0.475	0.618	0.480	0.564	0.097
GaussianNB	BorderlineSMOTE	0.598	0.501	0.618	0.480	0.672	0.136
SVC	TomekLinks	0.591	0.375	0.588	0.500	0.631	0.064
GaussianNB	ADASYN	0.576	0.453	0.588	0.462	0.679	0.140

Tabel 1 tampak bahwa model **DecisionTreeClassifier** dengan metode **sampling TomekLinks** memberikan performa terbaik berdasarkan metrik AUROC, AUPRC, dan akurasi. AUROC (Area Under the Receiver Operating Characteristic) adalah metrik yang umum digunakan untuk mengevaluasi performa model klasifikasi, dan nilai yang lebih tinggi menunjukkan performa yang lebih baik. Demikian pula, AUPRC (Area Under the Precision-Recall Curve) juga merupakan metrik penting, terutama dalam situasi di mana kelas tidak seimbang. Akurasi juga merupakan metrik yang umum digunakan, meskipun dapat menyesatkan jika kelas tidak seimbang. Disini juga terlihat nilai nilai F1 yang relatif rendah dibandingkan dengan beberapa model lainnya, dimana nilai ini merupakan nilai rata-rata harmonik dari presisi dan recall yang dapat digunakan sebagai metrik tunggal. Dalam table model ini terlihat nilai CV Mean (rata-rata cross-validation) yang relatif rendah dan CV Std (standar deviasi cross-validation) yang relatif tinggi. Ini dapat menunjukkan bahwa model ini mungkin memiliki variasi yang lebih besar dalam performanya saat diuji pada set data yang berbeda, yang dapat menunjukkan overfitting atau kurangnya generalisabilitas.

Diskusi

Kejadian infeksi kompleks *Entamoeba histolytica*/dispar/*moshkovskii* pada masyarakat pulau Weh masuk dalam katagori tinggi. **INI SEDANG ABI LANJUTKAN**

COBA SYIBBRAN BACA REVISIAN ABI, APA ADA KELIRU. KALAU ADA SYIBBRAN BENERIN DAN MUNGKIN KALAU MAU DI TAMBAH, TAMBAHKAN SAJA

Daftar pustaka

1. Shirley D, Hung C, Moonah S. Intestinal and Genital Infections [Internet]. *Entamoeba histolytica* (Amebiasis). Elsevier Inc.; 2015. 699–706 p. Available from: <https://doi.org/10.1016/B978-0-323-55512-8.00094-6>
2. Nasrallah J, Akhoundi M, Haouchine D, Marteau A, Mantelet S, Wind P, et al. Updates on the worldwide burden of amoebiasis: A case series and literature review. *J Infect Public Health* [Internet]. 2022;15(10):1134–41. Available from:

- <https://doi.org/10.1016/j.jiph.2022.08.013>
3. National Institutes of Health (NIH). NIAID Biodefense Research Agenda for Category B and C Priority Pathogens Progress Report. NIAID BIODEFENSE Prep Through Res [Internet]. 2004;1–72. Available from: https://www.niaid.nih.gov/sites/default/files/category_bc_progress_report.pdf
 4. Juriah N (Nully), Abdullah M (Murdani), Sutanto I (Inge), Chen K (Khie), Yuwono V (Vera). Intestinal Amebiasis: Diagnosis and Management. Indones J Gastroenterol Hepatol Dig Endosc [Internet]. 2005;6:80–5. Available from: <https://www.neliti.com/id/publications/67056/>
 5. Ishartadiati K. Protozoa and Bacteria Found on the Body of Flies at the Surabaya Market. ADOCPUB [Internet]. 2009; Available from: <https://adoc.pub/pendahuluan-bahan-dan-cara-kerja.html>
 6. Al-Areeqi MA, Sady H, Al-Mekhlafi HM, Anuar TS, Al-Adhroey AH, Atroosh WM, et al. First molecular epidemiology of *Entamoeba histolytica*, *E. dispar* and *E. moshkovskii* infections in Yemen: different species-specific associated risk factors. Trop Med Int Heal. 2017;22(4):493–504.
 7. van Hal, S. J., Stark, D. J., Fotedar, R., Marriott, D., Ellis, J. T., & Harkness JL. Amoebiasis: Current status in Australia. Med J Aust [Internet]. 2007;186(8):412–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/17437396/>
 8. Soares NM, Azevedo HC, Pacheco FTF, De Souza JN, Del-Rei RP, Teixeira MCA, et al. A Cross-Sectional Study of *Entamoeba histolytica*/*dispar*/*moshkovskii* Complex in Salvador, Bahia, Brazil. Biomed Res Int. 2019;2019.
 9. Directorate of Small Islands Empowerment. Weh [Internet]. Directorate of Small Islands Empowerment, Ministry of Maritime Affairs and Fisheries, Jakarta, Indonesia. 2017. p. 210093. Available from: http://www.ppk-kp3k.kkp.go.id/direktori-pulau/index.php/public_c/pulau_info/1142
 10. Gani A, Budiharsana MP. The Consolidated Report on Indonesia Health Sector Review 2018. Minist Natl Dev Plan Repub Indones. 2019;56.
 11. Birawida AB, Ibrahim E, Mallongi A, Rasyidi AA Al, Thamrin Y, Gunawan NA. Clean water supply vulnerability model for improving the quality of public health (environmental health perspective): A case in Spermonde islands, Makassar Indonesia. Gac Sanit [Internet]. 2021;35:S601–3. Available from: <https://doi.org/10.1016/j.gaceta.2021.10.095>
 12. Awad M, Khanna R. Efficient learning machine theories, concepts, and applications for engineers and system designers. *Apress*. 2015; 1-18.
 13. Mathews SM. Explainable Artificial Intelligence Applications in NLP, Biomedical, and Malware Classification: A Literature Review [Internet]. Vol. 998, Advances in Intelligent Systems and Computing. Springer International Publishing; 2019. 1269–1292 p. Available from: http://dx.doi.org/10.1007/978-3-030-22868-2_90
 14. Amin HA, Ali SA. Evaluation of different techniques of stool examination for intestinal parasitic infections in Sulaimani city - Iraq. Int J Curr Microbiol Appl Sci. 2015;4(5):991–6.
 15. Garcia LS, Arrowood M, Kokoskin E, Paltridge GP, Pillai DR, Procop GW, et al. Laboratory diagnosis of parasites from the gastrointestinal tract. Clin Microbiol Rev. 2018;31(1).
 16. ThermoFisher Scientific. Wheatley trichrome stain. ThermoFisher Sci [Internet]. 2021;1–2. Available from: <https://www.thermofisher.com/search/results?query=Wheatley+Trichrome+Stain&persona=DocSupport>
 17. Kesetyaningsih TW, Riswari RA, Pitaka RT. Prevalence Distribution of Intestinal Parasite Infestation in Under Five Years Children with Severe Malnutrition in Kasihan,

- Bantul, Yogyakarta Based on Risk Factors. *Mutiara Med.* 2010;10(2):135–41.
18. Sungkar S, Pohan APN, Ramadani A, Albar N, Azizah F, Nugraha ARA, et al. Heavy burden of intestinal parasite infections in Kalena Rongo village, a rural area in South West Sumba, eastern part of Indonesia: A cross sectional study. *BMC Public Health* [Internet]. 2015;15(1):1–6. Available from: <http://dx.doi.org/10.1186/s12889-015-2619-z>
 19. Sianturi MDG, Rahakbauw IM, Meyanti F, Kusumasari RA, Hartriyanti Y, Elsa Herdiana Murhandarwati E. Prevalence of intestinal protozoan infections and association with hygiene knowledge among primary schoolchildren in salahutu and Leihitu districts, central Maluku regency, Indonesia. *Trop Biomed.* 2016;33(3):428–36.