

KLASIFIKASI TEKS KLINIKAL BERBAGAI LABEL DENGAN MEKANISME PERHATIAN

LIM GUAN PENG, WANDEEP KAUR A/P Ratan Singh

^{1,2}Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,,
Selangor Darul Ehsan, Malaysia

Abstrak

Projek ini menilai keberkesanan *attention mechanism* dalam meningkatkan aplikasi dalam biomedis dan bioinformatik. Diilhamkan oleh perhatian kognitif, *attention mechanism* membolehkan model untuk menetapkan berat yang berbeza kepada bahagian-bahagian tertentu data input, meningkatkan tumpuan kepada istilah yang berkaitan dengan konteks. Walaupun model-model seperti GHS-NET telah maju dengan mengintegrasikan *convolutional neural network* (CNN) dan *recurrent neural network* (RNN), mereka mungkin masih gagal dalam memahami dan mengutamakan istilah penting untuk klasifikasi yang tepat. Kekurangan ini boleh menjejaskan prestasi mereka dalam memproses pelbagai teks bioperubatan yang terdapat dalam kesusasteraan dan nota klinikal. Kebaikan kajian ini adalah dalam mengintegrasikan *attention mechanism* untuk meningkatkan model pembelajaran yang mendalam dengan memberi penekanan pada teks yang relevan. Mekanisme ini mengira pemberat lembut untuk setiap perkataan, menandakan kepentingannya dalam konteks. Hasilnya, model itu menjadi mahir memfokuskan secara dinamik pada istilah bioperubatan yang penting, dengan itu memperhalusi pemahaman dan pemrosesan data inputnya. Menggabungkan *attention mechanism* ke dalam model pembelajaran mesin memajukan pengekstrakan pengetahuan, menghasilkan cerapan yang lebih tepat daripada kesusasteraan bioperubatan. Di luar akademik, kejayaan ini menjanjikan penjagaan kesihatan, meningkatkan penyelidikan perubatan dan amalan klinikal dengan memudahkan pengekstrakan pengetahuan unggul daripada data bioperubatan, yang berpotensi meningkatkan penjagaan pesakit dan memupuk kemajuan dalam sains perubatan.

Pengenalan

Dalam landskap penjagaan kesihatan yang sentiasa berkembang, klasifikasi teks telah muncul sebagai domain penting dan semakin jelas, terutamanya dalam bidang perubatan. Proses klasifikasi adalah langkah penting ke arah pembangunan aplikasi perubatan yang lebih praktikal. Pelbagai jenis data perubatan boleh dikategorikan kepada dua kumpulan yang berbeza, iaitu pengetahuan berstruktur dan tidak berstruktur. Kebanyakan nota teks perubatan sering tidak berstruktur yang menjadikannya sumber kesihatan yang berharga kerana ia mengandungi maklumat klinikal yang kaya mengenai keadaan pesakit (Lu et al. 2022). Maklumat yang diekstrak dari nota perubatan yang tidak berstruktur dengan klasifikasi teks dapat meningkatkan pemahaman pengamal perubatan dan peningkatan terhadap sistem penjagaan kesihatan. Klasifikasi teks perubatan tidak berstruktur juga dapat membantu dalam analisis dan pengambilan keputusan pengamal perubatan dengan menguraikan asal-usul komplikasi yang berkaitan dengan kesihatan. Pembelajaran automatik ontologi khusus domain yang fleksibel sepenuhnya dari teks yang tidak berstruktur masih merupakan kawasan yang kurang maju tetapi pastinya ia mempunyai potensi yang lebih besar untuk merevolusikan pelbagai tugas bioperubatan (Ibrahim et al. 2021).

Projek ini terutamanya memberi tumpuan kepada domain klasifikasi teks berbilang label dalam bidang klasifikasi teks perubatan. Klasifikasi teks perubatan juga boleh dikelaskan kepada dua jenis, klasifikasi pelbagai kelas dan pelbagai label di mana setiap dokumen bertulis dikaitkan dengan atau dikategorikan ke dalam satu label atau kelas dan dokumen mungkin mempunyai satu atau lebih label masing-masing. Walaupun kajian yang sedia ada kebanyakannya memberi tumpuan kepada menilai keberkesanan rangkaian saraf yang berbeza atau kaedah ensemble untuk mencapai prestasi optimum dalam klasifikasi teks perubatan (Ibrahim et al. 2021), perlu ditekankan bahawa penggabungan *attention mechanism* bersama dengan *Bidirectional Long Short-Term Memory* (Bi-LSTM) telah ditunjukkan untuk memberi kesan yang ketara pada bidang klasifikasi teks perubatan (Hui et al. 2020). *Attention mechanism* mengekstrak perkataan dan ayat yang bermakna, kemudian

mengumpulkan perwakilan ini untuk membentuk vektor ayat yang bermanfaat dalam menangkap ciri-ciri teks utama yang tidak dimiliki oleh kebanyakan model tradisional. Maklumat yang lebih penting, semakin tinggi berat yang diberikan, yang mengakibatkan pengaruh yang lebih besar pada klasifikasi. Pengiraan *attention mechanism* boleh dibahagikan kepada dua langkah. *Attention mechanism* bermula dengan mengira pengagihan perhatian dalam semua urutan input diikuti dengan pengiraan berat maklumat input mengikut pengagihan perhatian. Oleh itu, projek ini bertujuan untuk meneroka potensi *attention mechanism* dengan model yang ditentukan dalam konteks klasifikasi teks perubatan berbilang label, akhirnya menyumbang kepada matlamat yang lebih luas untuk meningkatkan ketepatan, kecekapan, dan kefahaman mengklasifikasikan teks perubatan dengan pelbagai label, dengan itu memajukan bidang informatika perubatan dan penjagaan pesakit. Kajian yang dijalankan menunjukkan bahawa model saraf yang dilengkapi dengan *attention mechanism* telah terbukti menunjukkan hasil yang menjanjikan (He et al. 2021). *Attention mechanism* mempunyai banyak jenis, termasuk berasaskan kandungan, berasaskan kedudukan dan lain-lain, dengan setiap jenis disesuaikan dengan tugas dan keperluan tertentu.

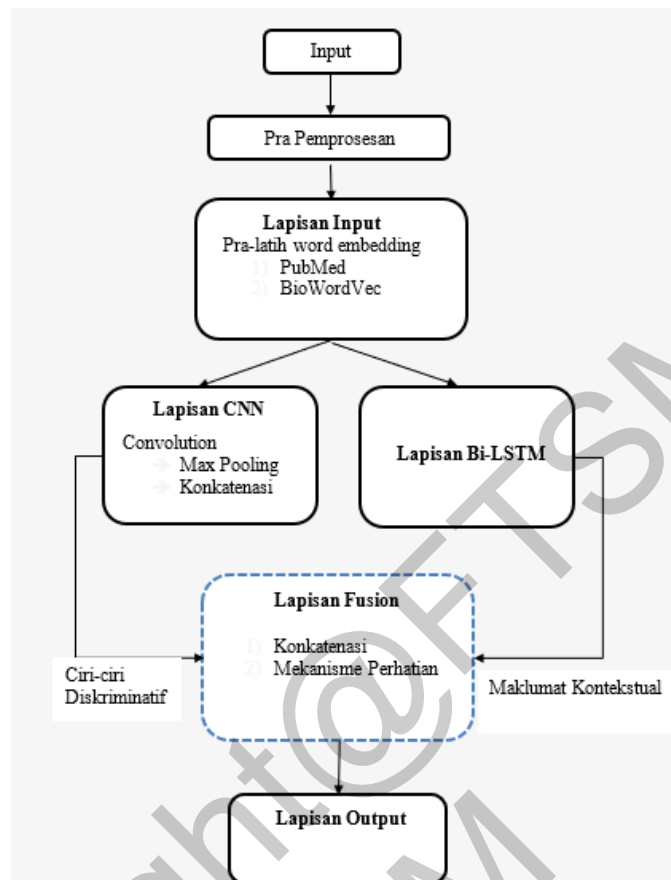
Objektif Kajian ini adalah seperti berikut:

Kajian Objektif

1. Untuk menyiasat domain klasifikasi teks perubatan pelbagai label
2. Untuk mencadangkan pelaksanaan model pembelajaran mendalam tertentu seperti BI-LSTM dan CNN serta *attention mechanism* yang membolehkan model untuk memberi tumpuan kepada terminologi perubatan penting atau teks dalam input.
3. Menilai prestasi model pada klasifikasi teks perubatan menggunakan metrik penilaian yang ditakrifkan.

Skop projek ini, "Klasifikasi teks klinikal berbagai label dengan mekanisme perhatian" akan memberi tumpuan kepada klasifikasi text perubatan dengan menambah *attention mechanism* ke dalam pangkalan data MIMIC-III, yang terdiri bukan sahaja daripada teks perubatan tetapi juga nota klinikal. Projek ini tidak akan termasuk pengklasifikasian imej perubatan tetapi hanya berdasarkan data berasaskan teks daripada pangkalan data MIMIC-III.

Metodologi Kajian



Rajah 1: Seni bina model yang dicadangkan untuk klasifikasi teks klinikal berbilang label

A. Pengumpulan Dan Pra Pemrosesan Input

Kajian ini menggunakan MIMIC III¹ (Medical Information Mart for Intensive Care), set data penanda aras yang diperolehi dari pangkalan data PhysioNet, yang merangkumi rekod perubatan individu (V et al. 2023). Dalam kajian ini, kami hanya menggunakan lajur TEKS dari data NOTEVENT. Teknik pra-proses yang akan dijalankan ialah penghapusan *stopword*, penghapusan whitespace, penghapusan nombor, penukaran huruf besar kepada huruf kecil, penapisan data yang kurang dari ambang dan penghapusan tanda baca. Pra-pemrosesan lain yang dilakukan ialah *one-hot encoding* pada lajur sasaran iaitu lajur kod ICD-9.

¹ <https://physionet.org/content/mimiciii/1.4/>

B. Kaedah Dan Analisis

Dalam lapisan input, *word embedding* yang telah dilatih dalam pangkalan data perubatan seperti BioWordVec dan Bio Clinical BERT digunakan. Dalam lapisan model, rangkaian saraf konvulsional (CNN) multichannel dengan dua saluran akan digunakan untuk mengekstrak ciri-ciri diskriminatif dari teks klinikal. Ini dicapai dengan memanfaatkan pra-latih *word embedding* dalam lapisan input. Bidirectional LSTM juga akan digunakan untuk memproses data berurutan secara dwihala, menangkap ketergantungan dalam kedua-dua arah ke hadapan dan ke belakang. LSTM menggunakan sel, pintu input, output, dan lupakan untuk memahami maklumat semantik dan menentukan maklumat apa yang harus diingat dan dilupakan (Ibrahim et al. 2021). Demikian juga dalam BI-LSTM, *dropout* juga dapat digunakan untuk mengurangkan risiko *overfitting* data semasa melatih model.

Attention mechanism diterapkan dalam lapisan penggabungan kerangka kerja ini. *Attention mechanism* secara dinamik memberikan nilai berat kepada kepentingan ciri-ciri yang digabungkan dari model yang berbeza (Ma et al. 2023). Ini dapat meningkatkan interpretabiliti input (Zhou et al. 2021) dan meningkatkan keupayaan model untuk memberi tumpuan kepada butiran yang relevan dalam teks input. Ini sangat bermanfaat dalam situasi di mana teks input adalah meluas, merangkumi pelbagai bahagian maklumat, memerlukan model untuk mengenali corak yang berkaitan untuk setiap label. Penggabungan *attention mechanism* ke dalam model-model seperti CNN dan RNN juga telah menunjukkan hasil yang positif dalam bidang klasifikasi teks (Ruan et al. 2022). Lapisan ini akan menciptakan representasi yang digabungkan yang mencakup ciri-ciri yang disesuaikan secara dinamik dari kedua-dua model pembelajaran mendalam. Kerangka kerja yang dicadangkan akan memanfaatkan fungsi aktivasi (sigmoid) yang menghasilkan kebarangkalian untuk setiap label dalam setiap teks. Label yang mempunyai kebarangkalian tertinggi akan dipilih sebagai label akhir bagi teks tertentu.

C. Penilaian

Pengukur dan alat ukur yang digunakan dalam projek ini adalah metrik penilaian seperti accuracy, ketepatan, ingatan, skor f1, AUC-ROC dan hamming loss yang dapat menilai kualiti prestasi model pembelajaran mesin.

Keputusan dan Perbincangan

A. Penilaian Penalaan Hiperparameter

Penalaan hiperparameter secara manual dijalankan untuk mendapatkan hasil prestasi model yang optimum. Penalaan hiperparameter pada kadar pembelajaran, set saiz penapis, dan dropout telah dilakukan untuk model CNN dan direkodkan dalam Jadual 1 dan 2, manakala penalaan hiperparameter pada saiz unit dan dropout telah dilakukan untuk model BI-LSTM dan direkodkan dalam Jadual 3 dan 4. Analisis menunjukkan bahawa model CNN dengan kombinasi saiz penapis 3 dan 5, bersama dengan kadar dropout 0.2, mencapai prestasi yang paling optimal. Oleh itu, tetapan ini digunakan dalam model CNN.

Sebaliknya, untuk model BI-LSTM, analisis menunjukkan bahawa kadar pembelajaran 0.002, gabungan unit tersembunyi 128 dan 2 lapisan bersama dengan kadar penurunan 0.5 terbukti menjadi pilihan yang paling optimal, mencapai prestasi terbaik di semua ukuran penilaian. Oleh itu, tetapan ini digunakan dalam model BI-LSTM.

Model	Learning Rate	Precision	Recall	f1-score	Accuracy	Hamming Loss
CNN	0.001	0.77	0.53	0.63	0.83	0.17
BI-LSTM	0.001	0.78	0.55	0.65	0.84	0.16
CNN+BI-LSTM	0.001	0.26	0.56	0.36	0.49	0.51
CNN	0.002	0.75	0.50	0.60	0.82	0.18
BI-LSTM	0.002	0.77	0.60	0.67	0.84	0.16
CNN+BI-LSTM	0.002	0.25	0.54	0.34	0.46	0.54
CNN	0.003	0.76	0.49	0.60	0.82	0.18
BI-LSTM	0.003	0.73	0.58	0.65	0.83	0.17
CNN+BI-LSTM	0.003	0.25	0.53	0.34	0.48	0.52
CNN	0.004	0.76	0.50	0.60	0.83	0.17
BI-LSTM	0.004	0.73	0.61	0.66	0.83	0.17
CNN+BI-LSTM	0.004	0.25	0.49	0.33	0.50	0.50
CNN	0.005	0.74	0.54	0.62	0.83	0.17
BI-LSTM	0.005	0.74	0.60	0.67	0.84	0.16
CNN+BI-LSTM	0.005	0.26	0.53	0.35	0.48	0.52

Jadual 1: Eksperimen Kadar Pembelajaran untuk Model

Saiz Penapis	Ketepatan	Ingatan Semula	Skor F1	Accuracy	Hamming Loss
3,4	0.75	0.54	0.63	0.83	0.17
3,5	0.77	0.53	0.63	0.83	0.17
4,3	0.76	0.52	0.61	0.83	0.17
4,5	0.76	0.52	0.62	0.83	0.17

Jadual 2 Penalaan Hiperparameter untuk Model CNN

Dropout	Ketepatan	Ingatan Semula	Skor F1	Accuracy	Hamming Loss
0.1	0.76	0.58	0.66	0.84	0.16
0.2	0.78	0.52	0.60	0.83	0.17
0.3	0.76	0.51	0.61	0.83	0.17
0.4	0.76	0.51	0.61	0.83	0.17
0.5	0.77	0.53	0.63	0.83	0.17

Jadual 3 Penalaan Hiperparameter untuk Model CNN

Bilangan Unit	Bilangan Lapisan	Ketepatan	Ingatan Semula	Skor F1	Accuracy	Hamming Loss
64	3	0.72	0.42	0.53	0.80	0.20
64	2	0.73	0.52	0.61	0.82	0.18
128	3	0.71	0.54	0.62	0.82	0.18
128	2	0.77	0.60	0.67	0.84	0.16

Jadual 4 Penalaan Hiperparameter untuk Model BI-LSTM

Dropout	Precision	Recall	f1-score	Accuracy	Hamming Loss
0.1	0.75	0.54	0.63	0.83	0.17
0.2	0.76	0.49	0.60	0.82	0.18
0.3	0.77	0.51	0.62	0.83	0.17
0.4	0.78	0.51	0.62	0.83	0.17
0.5	0.77	0.60	0.67	0.84	0.16

Jadual 5 Penalaan Hiperparameter untuk Model BI-LSTM

B. Prestasi Keseluruhan Model

Dalam eksperimen ini, model CNN, BI-LSTM, dan model ensemble kedua-dua rangkaian tersebut digunakan dalam klasifikasi teks klinikal dengan pelbagai label. Berdasarkan Jadual 6, dapat diperhatikan bahawa model BI-LSTM dengan mekanisme perhatian mencatatkan prestasi yang lebih tinggi berbanding model CNN dengan perhatian. Prestasi unggul model BI-LSTM berbanding CNN mungkin disebabkan oleh keupayaan BI-LSTM untuk mengekstrak maklumat konteks dengan menggunakan ciri-ciri memori dan pelupusan mekanisme gerbang, yang bukan sahaja boleh menangkap korelasi urutan panjang dengan berkesan, tetapi juga mengurangkan masalah *vanishing gradient*. Sebaliknya, CNN tidak dapat menangkap ciri jarak jauh (Wang et al. 2022). Bukan itu sahaja, BI-LSTM dengan *self-attention* juga berprestasi lebih baik dengan *accuracy* yang lebih tinggi dan kerugian hamming yang lebih rendah berbanding dengan model lain.

	Mikro	Makro	Weighted	Sampel
CNN dengan Attention	0.63	0.58	0.61	0.57
BI-LSTM dengan Attention	0.65	0.59	0.63	0.61
CNN+BI-LSTM dengan Attention	0.33	0.32	0.34	0.30

Jadual 6: Perbandingan skor f1 antara model dengan self-attention

	Accuracy	Hamming Loss
CNN dengan Attention	0.83	0.17
BI-LSTM dengan Attention	0.84	0.16
CNN+BI-LSTM dengan Attention	0.51	0.48

Jadual 7: Perbandingan accuracy dan hamming loss model dengan self-attention

C. Prestasi Keseluruhan Model Berdasarkan Label Individu

Perbandingan skor f1 bagi setiap label 10 ICD-9 teratas terus dijalankan dan direkodkan dalam Jadual 7 untuk membezakan perbezaan model dengan lebih baik. Berdasarkan Jadual 4.6.2, dapat diperhatikan bahawa ketiga-tiga model dengan *self-attention* mencatatkan peningkatan dalam

meramalkan label ICD-9 seperti 4280, 5849 dan 53081 berdasarkan peningkatan skor f1. Dalam kajian saya terkini, peningkatan pada label-label tertentu disebabkan oleh kehadiran terminologi yang tepat dan jelas. Selain itu, hipotesis saya mengenai sebab mengapa self-attention bertambah baik untuk kod ICD-9 4280 (kegagalan jantung kongestif) berbanding dengan 42731 (atrial fibrillation) adalah kerana mekanisme *self-attention* mungkin mengabaikan singkatan dan terminologi penting seperti "associated with AF" dalam sejarah penyakit semasa. Perbezaan halus ini dapat memberi pengaruh signifikan terhadap *weightage* yang diberikan oleh mekanisme *self-attention*, dengan demikian mempengaruhi prestasinya.

Label	Judul Pendek	CNN	CNN (A)	BI- LSTM	BI- LSTM(A)	CNN+BI- LSTM	CNN+BI- LSTM(A)
4019	Hypertension NOS	0.67	0.63	0.66	0.66	0.42	0.48
4280	Congestive heart failure NOS	0.71	0.72	0.73	0.71	0.32	0.43
42731	Atrial fibrillation	0.88	0.87	0.87	0.87	0.38	0.37
41401	Cornary atherosclerosis native vessel	0.69	0.63	0.73	0.75	0.43	0.39
5849	Acute kidney failure NOS	0.56	0.60	0.54	0.59	0.29	0.31
25000	Diabetes mellitus without complication unspecified	0.30	0.19	0.32	0.39	0.21	0.30
2724	Hyperlipidemia NEC/NOS	0.61	0.56	0.40	0.59	0.32	0.28
51881	Acute respiratory failure	0.51	0.52	0.53	0.55	0.32	0.30
5990	Urinary tract infection NOS	0.40	0.35	0.36	0.23	0.28	0.25
53081	Esophageal reflux	0.65	0.72	0.27	0.61	0.21	0.25

Jadual 7 : Perbandingan Skor F1 bagi Kod ICD-9 Individu

Seterusnya, penilaian AUC-ROC berdasarkan 10 label ICD-9 teratas bagi setiap model dengan dan tanpa pelaksanaan mekanisme perhatian sendiri dijalankan dan direkodkan dalam Jadual 8. Berdasarkan jadual ini, jelas bahawa AUC-ROC untuk setiap label kelas melebihi 0.5, menunjukkan tahap keupayaan ramalan dalam kedua-dua model. Tambahan pula, adalah ketara

bahawa dalam kebanyakan label kelas, model dengan pelaksanaan perhatian sendiri mencapai nilai AUC-ROC yang lebih tinggi berbanding model tanpa pelaksanaan perhatian sendiri.

Label	Judul Pendek	CNN	CNN (A)	BI-LSTM	BI-LSTM (A)	CNN+BI- LSTM	CNN+BI- LSTM (A)
4019	Hypertension NOS	0.76	0.65	0.71	0.78	0.50	0.55
4280	Congestive heart failure NOS	0.85	0.81	0.87	0.87	0.50	0.50
42731	Atrial fibrillation	0.94	0.90	0.95	0.95	0.49	0.55
41401	Crnry athrsc1 natve vssl	0.88	0.74	0.91	0.91	0.54	0.52
5849	Acute kidney failure NOS	0.83	0.71	0.81	0.84	0.50	0.48
25000	DMII wo cmp nt st uncnt	0.84	0.56	0.83	0.82	0.51	0.48
2724	Hyperlipidemia NEC/NOS	0.90	0.58	0.84	0.89	0.57	0.48
51881	Acute respiratory failure	0.89	0.68	0.90	0.89	0.47	0.52
5990	Urinary tract infection NOS	0.81	0.69	0.71	0.75	0.47	0.52
53081	Esophageal reflux	0.94	0.68	0.91	0.93	0.52	0.56

Jadual 8: Penilaian AUC-ROC

Kesimpulan

Secara rumusan, projek ini menyiasat klasifikasi teks klinikal berbilang label menggunakan mekanisme perhatian dengan dataset Medical Information Mart for Intensive Care III (MIMIC-III), sebuah pangkalan data komprehensif yang mengandungi rekod kesihatan tanpa identiti daripada pesakit penjagaan kritikal. Pelbagai model pembelajaran mendalam, termasuk rangkaian neural konvolusi (CNN), *bidirectional long short-term memory network* (Bi-LSTM), dan model ensemble yang menggabungkan CNN dan Bi-LSTM, digunakan dengan pengintegrasian mekanisme *self-attention* untuk meningkatkan klasifikasi teks klinikal berbilang label. Penalaan hiperparameter juga vital untuk mencapai prestasi model yang optimum. Penemuan utama menunjukkan bahawa mekanisme *self-attention* secara signifikan meningkatkan prestasi model, membuktikan kegunaannya. Antara model pembelajaran mendalam yang digunakan, Bi-LSTM dengan *self-attention* menunjukkan prestasi tertinggi, diikuti oleh CNN tunggal dengan *self-attention*, manakala model ensemble dengan *self-attention* menunjukkan prestasi terendah.

Memandangkan kepelbagaian mekanisme perhatian yang tersedia, penyelidikan masa depan boleh meneroka jenis tambahan untuk meningkatkan prestasi model. Contohnya termasuk perhatian *multi-head*, yang memanfaatkan pengiraan selari untuk memperkayakan model, perhatian *additive* yang menggunakan rangkaian neural untuk hubungan yang rumit, dan perhatian *multiplicative* menggunakan hasil titik untuk kecekapan. Setiap jenis menawarkan kekuatan unik yang boleh dioptimumkan untuk memperkukuhkan fokus model pada maklumat yang berkaitan, dengan itu meningkatkan prestasi merentas aplikasi yang pelbagai. Penerokaan mekanisme perhatian alternatif mungkin menghasilkan peningkatan yang ketara dalam prestasi model pada data seperti MIMIC-III. Selain itu, memandangkan ketidakseimbangan yang terdapat dalam set data MIMIC-III, penggunaan teknik seperti Teknik Pembahagian Minoriti Sintetik (SMOTE) boleh menjadi bermanfaat. SMOTE membantu menangani ketidakseimbangan kelas dengan menjana sampel sintetik untuk kelas minoriti, dengan itu meningkatkan keupayaan model untuk belajar daripada data yang kurang diwakili dan memperbaiki prestasi pengelasan keseluruhan.

Penghargaan

Pertama sekali, saya ingin mengucapkan terima kasih kepada Tuhan yang Maha Esa atas limpahan rahmat dan petunjuk-Nya sepanjang proses penyusunan tesis ini. Kepada penyelia Dr Wandeep Kaur a/p Ratan Singh, terima kasih atas bimbingan, sokongan, dan nasihat yang berharga sepanjang perjalanan penyelidikan ini.

Saya juga ingin menyampaikan penghargaan kepada Fakulti / Institut / Pusat Pengajian / Jabatan atas kemudahan penyelidikan yang disediakan, yang telah menyokong kelancaran laporan usulan ini.

Akhir sekali, saya juga mengucapkan ribuan terima kasih kepada semua pihak yang turut serta memberi sumbangan dan sokongan, yang telah menjadikan laporan usulan ini berjaya.

Saya tidak ketinggalan untuk berterima kasih kepada rakan-rakan dan keluarga saya yang telah berkerjasama dengan saya menjalankan laporan usulan saya. Mereka turut memberi bantuan dan bimbingan.

RUJUKAN

Alsentzer, E., Murphy, J., Boag, W., Weng, W.-H., Jin, D., Naumann, T. & Mcdermott, M. 2019.

Publicly Available Clinical BERT Embeddings.

Bahdanau, D., Cho, K. & Bengio, Y. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. ArXiv 1409:

Blanco, A., Remmer, S., Pérez, A., Dalianis, H. & Casillas, A. 2022. Implementation of specialised attention mechanisms: ICD-10 classification of Gastrointestinal discharge summaries in English, Spanish and Swedish. *Journal of Biomedical Informatics* 130: 104050.

Chen, P., Dong, W., Wang, J., Lu, X., Kaymak, U. & Huang, Z. 2020. Interpretable clinical prediction via attention-based neural network. *BMC Medical Informatics and Decision Making* 20(3): 131.

- Han, S., Zhang, R. F., Shi, L., Richie, R., Liu, H., Tseng, A., Quan, W., Ryan, N., Brent, D. & Tsui, F. R. 2022. Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. *Journal of Biomedical Informatics* 127: 103984.
- Hashir, M. & Sawhney, R. 2020. Towards unstructured mortality prediction with free-text clinical notes. *Journal of Biomedical Informatics* 108: 103489.
- He, W., Wu, Y. & Li, X. 2021. Attention Mechanism for Neural Machine Translation: A survey. 2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp.1485-1489.
- Hui, Y., Du, L., Lin, S., Qu, Y. & Cao, D. 2020. Extraction and Classification of TCM Medical Records Based on BERT and Bi-LSTM With Attention Mechanism. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp.1626-1631.
- Ibrahim, M. A., Ghani Khan, M. U., Mehmood, F., Asim, M. N. & Mahmood, W. 2021. GHS-NET a generic hybridized shallow neural network for multi-label biomedical text classification. *Journal of Biomedical Informatics* 116: 103699.
- Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L.-W. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L. & Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3(1): 160035.
- Khattak, F. K., Jeblee, S., Pou-Prom, C., Abdalla, M., Meaney, C. & Rudzicz, F. 2019. A survey of word embeddings for clinical text. *Journal of Biomedical Informatics* 100: 100057.
- Lentzas, A., Dalagdi, E. & Vrakas, D. 2022. Multilabel Classification Methods for Human Activity Recognition: A Comparison of Algorithms. *Sensors* 22(6): 2353.
- Lu, H., Ehwerhemuepha, L. & Rakovski, C. 2022. A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. *BMC Medical Research Methodology* 22(1): 181.

- Ma, Y., Huang, Z., Su, J., Shi, H., Wang, D., Jia, S. & Li, W. 2023. A Multi-Channel Feature Fusion CNN-Bi-LSTM Epilepsy EEG Classification and Prediction Model Based on Attention Mechanism. *IEEE Access* 11: 62855-62864.
- Makohon, I. & Li, Y. 2021. Multi-Label Classification of ICD-10 Coding & Clinical Notes Using MIMIC & CodiEsp. 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), pp.1-4.
- Maxwell, A., Li, R., Yang, B., Weng, H., Ou, A., Hong, H., Zhou, Z., Gong, P. & Zhang, C. 2017. Deep learning architectures for multi-label classification of intelligent health risk prediction. *BMC Bioinformatics* 18(14): 523.
- Nam, J., Choi, J. W., Shin, Y. G. & Park, S. 2023. A BERT-Based Artificial Intelligence to Analyze Free-Text Clinical Notes for Binary Classification in Papillary Thyroid Carcinoma Recurrence. 2023 IEEE International Conference on Consumer Electronics (ICCE), pp.1-2.
- Natsiavas, P., Rasmussen, J., Voss-Knude, M., Votis, K., Coppolino, L., Campegiani, P., Cano, I., Marí, D., Faiella, G., Clemente, F., Nalin, M., Grivas, E., Stan, O., Gelenbe, E., Dumortier, J., Petersen, J., Tzovaras, D., Romano, L., Komnios, I. & Koutkias, V. 2018. Comprehensive user requirements engineering methodology for secure and interoperable health data exchange. *BMC Medical Informatics and Decision Making* 18(1): 85.
- Nuthakki, S., Neela, S., Gichoya, J. & Purkayastha, S. 2019. Natural language processing of MIMIC-III clinical notes for identifying diagnosis and procedures with neural networks.
- Opalka, S., Stasiak, B., Szajerman, D. & Wojciechowski, A. 2018. Multi-Channel Convolutional Neural Networks Architecture Feeding for Effective EEG Mental Tasks Classification. *Sensors (Basel)* 18(10):
- Oubenali, N., Messaoud, S., Filiot, A., Lamer, A. & Andrey, P. 2022. Visualization of medical concepts represented using word embeddings: a scoping review. *BMC Medical Informatics and Decision Making* 22(1): 83.

- Prabhakar, A., Shidharth, S. & Kamath, S. S. 2022. Neural Language Modeling of Unstructured Clinical Notes for Automated Patient Phenotyping. 2022 56th Annual Conference on Information Sciences and Systems (CISS), pp.142-147.
- Prabhakar, S. K. & Won, D. O. 2021. Medical Text Classification Using Hybrid Deep Learning Models with Multihead Attention. *Comput Intell Neurosci* 2021: 9425655.
- Sen, C., Hartvigsen, T., Kong, X. & Rundensteiner, E. 2019. Patient-level Classification on Clinical Note Sequences Guided by Attributed Hierarchical Attention. 2019 IEEE International Conference on Big Data (Big Data), pp.930-939.
- V, V., S, S., P. K, R. & B, S. 2023. MIMIC III Text classification with the generalization of BERT transformer model synergized with XGBoost classifier. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp.1-6.
- Venkataraman, G. R., Pineda, A. L., Bear Don't Walk Iv, O. J., Zehnder, A. M., Ayyar, S., Page, R. L., Bustamante, C. D. & Rivas, M. A. 2020. FasTag: Automatic text classification of unstructured medical narratives. *PLoS One* 15(6): e0234647.
- Wang, S., Yang, Y. & Meng, X. 2022. Research on Multi-Label Text Classification Based on Multi-Channel CNN and BiLSTM. 2022 International Conference on Artificial Intelligence of Things and Crowdsensing (AIoTCs), pp.498-503.
- Wu, J., Huang, C. & Chen, Y. 2020. Patent Text Classification Study Based on Bi-LSTM-A Model. 2020 5th International Conference on Control, Robotics and Cybernetics (CRC), pp.1-5.
- Wu, Y., Zeng, M., Yu, Y., Li, Y. & Li, M. 2022. A Pseudo Label-Wise Attention Network for Automatic ICD Coding. *IEEE Journal of Biomedical and Health Informatics* 26(10): 5201-5212.
- Zhang, Y., Chen, Q., Yang, Z., Lin, H. & Lu, Z. 2019. BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Scientific Data* 6(1): 52.

- Zhang, Y., Li, X., Liu, Y., Li, A., Yang, X. & Tang, X. 2023. A Multilabel Text Classifier of Cancer Literature at the Publication Level: Methods Study of Medical Text Classification. *JMIR Med Inform* 11: e44892.
- Zhu, M., Ahuja, A., Wei, W. & Reddy, C. K. 2019. A Hierarchical Attention Retrieval Model for Healthcare Question Answering. The World Wide Web Conference, Organized by San Francisco, CA, USA. Association for Computing Machinery.
- Johnson, A., Pollard, T., & Mark, R. (2019). MIMIC-III Clinical Database Demo (version 1.4). PhysioNet.
- Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3, 160035.
- Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220.
- Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4): 772-780.

LIM GUAN PENG (A186683)
WANDEEP KAUR A/P Ratan Singh
Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia