KAJIAN PEMBELAJARAN MENDALAM ALAM PENGESANAN PANCINGAN DATA URL

Nur Syazwana binti Mohd Yunan, Wan Fariza binti Paizi@Fauzi

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

ABSTRAK

Kajian ini menangani masalah berterusan serangan pancingan data URL dengan memastikan keberkesanan pembelajaran mendalam dipertingkatkan dengan penyelarasan kecicirian dan pemilihan ciri. Matlamatnya ialah untuk menyiasat keberkesanan pembelajaran mendalam dalam mengesan URL pancingan data dari segi keteguhan dan kecekapan. Menggunakan pancingan data berlabel yang tersedia secara terbuka dan set data URL sah berlabel, kajian ini menjalankan eksperimen terkawal dimana perlaksanaan dan analisis model CNN 1D sedia ada sebagai model garis dasar. Model garis dasar ini mencapai ketepatan tinggi 0.91. Walau bagaimanapun, pada penyiasatan lanjut, didapati model tersebut mengalami masalah pemasangan berlebihan jika dilihat daripada graf kehilangan dan ketepatan. Kami melihat pada regularisasi dan pemilihan ciri untuk menangani isu pemasangan berlebihan. Penemuan utama menunjukkan bahawa ciri sederhana secara berkesan menstabilkan prestasi pengesahan manakala pemilihan ciri mengekalkan kadar pengesanan 0.88 sambil mengurangkan perbezaan antara kehilangan latihan dan pengesahan. Sumbangan kajian ini termasuk reka bentuk saluran paip CNN dengan ciri pemilihan yang menyediakan panduan praktikal untuk mengurangkan pemasangan berlebihan melalui nilai kecicirian yang sesuai dan pengurangan dimensi, dengan itu meningkatkan keberkesanan dan kebolehoperasian pengesanan pancingan data.

Kata kunci: pembelajaran mendalam, pancingan data, pemasangan berlebihan, nilai kecicirian, pemilihan ciri

1.0 PENGENALAN

Pancingan data merujuk kepada jenayah siber iaitu percubaan untuk mencuri maklumat sensitif untuk menipu seseorang bagi mendapatkan maklumat peribadi dan memindahkan wang. Maklumat peribadi ini kemudiannya digunakan oleh penyerang untuk merugikan mangsa. Istilah pancingan data berasal daripada perkataan “memancing”. Logik terminologi ini ialah penyerang menggunakan “umpan” untuk menarik minat mangsa dan kemudian “memancing” maklumat peribadi mereka. Di era yang semakin maju seperti sekarang ini, pancingan data URL sering diedarkan melalui e-mel, sosial mesej media, atau platform dalam talian lain, disertai oleh alasan yang meyakinkan direka untuk menarik mangsa untuk klik pada pautan yang berniat jahat (Butnaru et al., 2021).

Serangan pancingan data dimana mangsa menerima mesej yang telah dihantar oleh kenalan atau organisasi yang dikenali. Mesej itu mengandungi perisian hasad yang

menyasarkan pengguna komputer atau mempunyai pautan untuk mengarahkan mangsa ke laman web berniat jahat atau palsu untuk menipu mereka supaya mendedahkan maklumat peribadi dan kewangan, seperti kata laluan, ID akaun atau butiran kad kredit. Pancingan data URL melibatkan penciptaan pautan web yang mengelirukan yang meniru URL yang sah untuk menipu pengguna supaya mereka mengakses laman web atau perkhidmatan yang dipercayainya (Albishri & Dessouky, 2024).

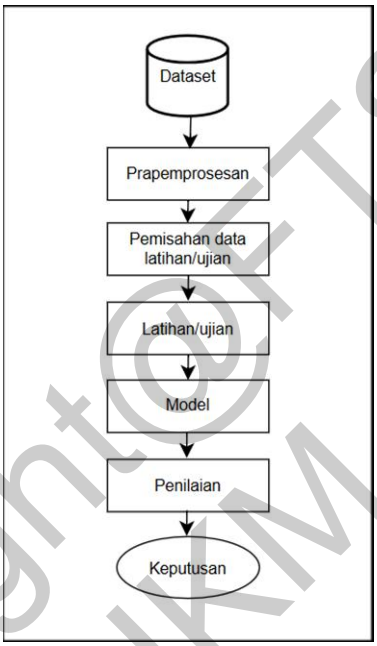
Terdapat banyak aplikasi pengesanan yang telah dibangunkan. Namun, aplikasi pengesanan pancingan data yang sedia ada ini terdapat beberapa kelemahannya. Banyak kaedah pengesanan pancingan data termasuk sistem berasaskan heuristik dan beberapa model pembelajaran mesin mengalami kesukaran dengan keseimbangan antara positif palsu iaitu salah menanda URL adalah sah pancingan data dan negatif palsu iaitu gagal mengesan URL pancingan data yang sebenar (Opara, Chen & Wei 2023) . Sistem ini boleh dipintas oleh penyerang yang sentiasa mengembangkan taktik-taktik mereka untuk menipu. Terdapat keperluan untuk model sistem pengesanan yang ditambah baik supaya dapat mengatasi ancaman baharu dan memberikan perlindungan yang menyeluruh.

Pembelajaran mesin dan teknik pembelajaran mendalam sedang dilaksanakan untuk meningkatkan keselamatan siber. Kecerdasan buatan (AI) ialah kaedah risikan yang berkembang pesat yang membantu mengawasi keselamatan siber dan menyediakan perlindungan untuk aktiviti komputer (S, A., & M. S., V. P. 2024). . Prestasi model pengesanan pancingan data berasaskan pembelajaran mesin (ML) tradisional semakin merosot dari masa ke semasa. Kegagalan ini disebabkan oleh perubahan drastik dalam pengedaran ciri yang disebabkan oleh teknik pancingan data baharu dan evolusi teknologi dari masa ke semasa (Ejaz et al., 2023). Walau bagaimanapun, kajian menunjukkan bahawa model pembelajaran mendalam CNN 1D mencapai ketepatan latihan yang tinggi cenderung mengalami masalah pemasangan berlebihan iaitu ketidakupayaan membuat generalisasi kepada data baharu (Yerima et al., 2020).

Oleh hal yang demikian, kajian ini dilaksanakan untuk menambah baik dan menggunakan teknologi yang terkini iaitu menyiasat model pembelajaran mendalam yang sedia ada untuk pengesanan pancingan data URL berkaitan pemasangan berlebihan, membandingkan model pembelajaran mendalam CNN 1D dengan kadar kecicirán dan menyiasat pemilihan ciri dalam menangani isu pemasangan berlebihan. Pembelajaran dalam adalah salah satu kaedah kecerdasan buatan (AI) yang merupakan bidang ilmu pengetahuan terkini yang digunakan secara meluas. Untuk menangani ancaman berterusan serangan pancingan data, teguh, langkah keselamatan siber diperlukan telah muncul sebagai pendekatan yang menjanjikan. Kemajuan dalam pembelajaran mendalam telah membuat kemungkinan mesin untuk belajar sendiri dan mengekstrak ciri secara automatik tanpa campur tangan manusia menjadikan keseluruhan proses lebih cekap dan kos efektif. Pembelajaran mendalam telah merevolusikan cara untuk mendekati masalah yang kompleks dengan menyediakan lebih cepat dan penyelesaian yang lebih tepat, terutamanya semasa menangani set data dan tugasan yang besar dalam pengesanan pancingan data (S, A., & M. S., V. P. 2024).

2.0 METODOLOGI KAJIAN

Model ini adalah model konsep yang digunakan untuk menerangkan keadaan suatu sistem. Ianya juga merupakan satu gambaran atau pelan yang digunakan untuk merancang dan memahami model yang digunakan ini, serta untuk membuat keputusan tentang bagaimana model ini diimplementasi, diurus, dan ditambah baik. Saluran paip model yang menunjukkan rangka kerja umum di mana pengesanan merangkumi pengumpulan data, pra-pemprosesan, latihan model, penilaian dan ramalan seperti yang ditunjukkan di rajah 1.



Rajah 1 Model Saluran Paip

2.1 DATASET

Langkah awal dalam membangunkan sistem pengesanan pancingan data ialah mengumpul set data yang teguh dan didapatkan dari sumber-sumber yang sah, yang sepatutnya mengandungi pancingan data dan URL yang sah. Set data yang digunakan dalam percubaan ini ialah "Set Data Pengesanan Phishing Halaman Web" (2021) daripada (Web Page Phishing Detection Dataset, 2021) yang juga digunakan oleh Bakare-Opeyemi (n.d.) . Ia terdiri daripada 11,430 sampel dengan contoh pancingan data dan tapak web yang sah yang sama iaitu 5715. Setiap sampel termasuk URL dan 87 ciri berkaitan yang menangkap pelbagai sifat halaman web/URL, bersama-sama dengan label (status) yang menunjukkan sama ada tapak pancingan data (1) atau sah (0).

2.2 PRA PEMPROSESAN

Peringkat seterusnya ialah Pra Pemprosesan, di mana data URL yang tidak diproses diubah menjadi input berstruktur yang boleh digunakan oleh model pembelajaran mendalam. Langkah ini termasuk tokenisasi, vektorisasi dan normalisasi. Tokenisasi bermaksud Setiap URL dipecahkan kepada sub-komponennya, seperti nama domain, laluan dan rentetan pertanyaan. Di mana vektorisasi pula melibatkan transformasi token kepada perwakilan berangka, seperti benam, yang merangkum sambungan semantik. Selain itu, normalisasi pula bermaksud konsistensi dengan menyeragamkan panjang URL melalui padding atau pemotongan.

Langkah-langkah yang telah kami lakukan dalam fasa ini adalah melibatkan pra pemrosesan komprehensif set data untuk memastikan ia bersih, seimbang dan berstruktur dengan betul untuk melatih model. Pada mulanya, set data diimport menggunakan pandas.read_csv() dan diperiksa menggunakan fungsi seperti head() untuk memahami strukturnya. Berikutan ini, set data disahkan untuk baki kelas. Dalam kes ini, ia mengandungi bilangan pancingan data dan URL sah yang sama, yang sesuai iaitu 5715 dimana jika ditambah kedua dua kategori berjumlah 11430 url.

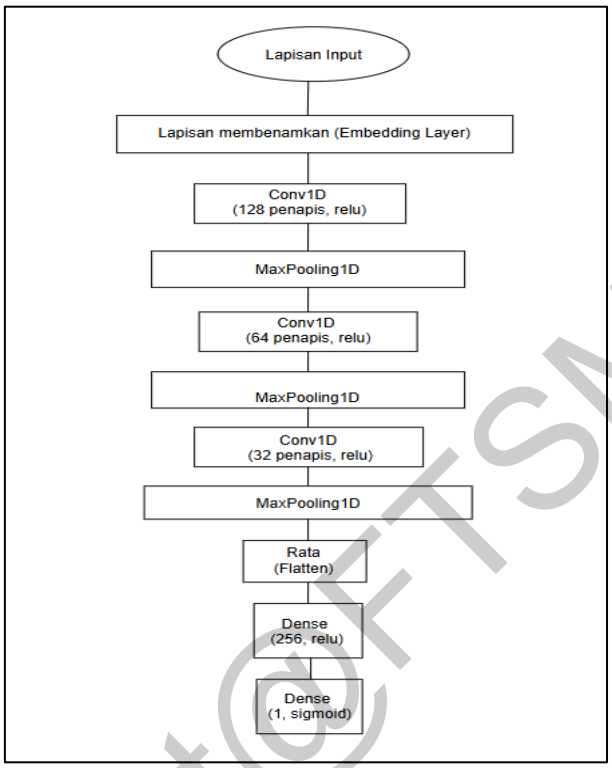
2.3 PEMISAHAN DATA UJIAN DAN LATIHAN

Set data dibahagikan kepada subset latihan dan ujian menggunakan train_test_split, dengan perkadaran 80% untuk latihan dan 20% untuk ujian yang memberikan 9144 sampel untuk data latihan dan 2286 sampel untuk data ujian. Stratifikasi memastikan bahawa kedua-dua subset mengekalkan perkadaran yang sama bagi pancingan data dan sampel yang sah, yang penting untuk penilaian prestasi model yang seimbang.

2.4 LATIHAN DAN UJIAN

Kedua-dua model garis dasar CNN dan model CNN dengan pemilihan ciri dilatih di bawah tetapan setanding untuk menilai prestasi mereka. Proses latihan telah dijalankan pada persediaan perkakasan standard dengan menggunakan GPU dalam Google Colab. Pengoptimum Adam digunakan dengan kadar pembelajaran 0.001. Fungsi kehilangan ialah entropi silang binari, yang sesuai untuk klasifikasi dua kelas dan berfungsi dengan baik dengan output sigmoid. Semasa latihan, kehilangan/ketepatan latihan dan kehilangan/ketepatan pengesahan dipantau. Setiap model dinilai sama ada ciri tambahan dalam gabungan model CNN dengan pemilihan ciri membantu untuk belajar lebih cepat atau mencapai ekspektasi . Untuk setiap latihan, metrik direkodkan untuk membandingkan keputusan pada set ujian kemudian.

2.5 MODEL CNN 1D



Rajah 2 CNN 1D

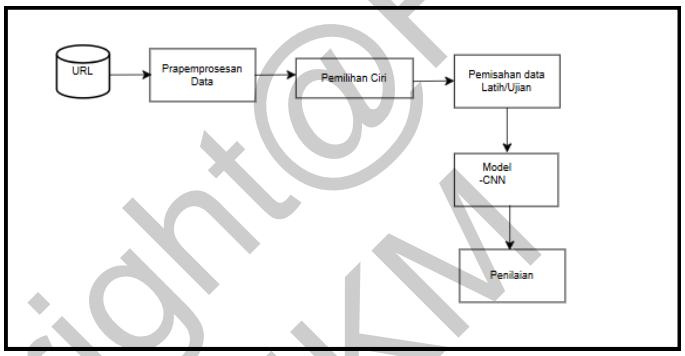
Model pengesanan pancingan data ini terdiri daripada Rangkaian saraf Konvolusi (CNN), iaitu rangkaian saraf tiruan yang mendalam, ia mengandungi lapisan input, tiga lapisan lilitan 1D, tiga lapisan pengumpulan maksima 1D, lapisan rata dan dua lapisan padat seperti yang ditunjukkan dalam rajah 2. Setiap lapisan mempunyai peranan yang berbeza untuk tujuan tertentu dalam pengekstrakan ciri dan pembuatan keputusan. Dalam seni bina ini, model bermula dengan lapisan membenam, yang biasa digunakan dalam CNN berasaskan teks. Lapisan ini menukar indeks kata input menjadi vektor padat 300-dimensi, membolehkan model memahami hubungan semantik diantara setiap satu perkataan.

Lapisan utama pertama ialah Lapisan Konvolusi (Conv1D). Lapisan ini menggunakan 128 penapis dan saiz kernel 4 untuk mengimbas input yang terbenam. Lapisan ini bertanggungjawab untuk mengesan ciri-ciri tempatan dalam teks seperti corak 4 perkataan berturut-turut. Fungsi pengaktifan yang digunakan adalah ReLU (Rectified Linear Unit), yang memperkenalkan ketidaklinieran dengan hanya mengekalkan nilai positif di mana ia dapat membantu model mempelajari hubungan kompleks. Seterusnya, lapisan Max Pooling 1D dengan saiz kolam 2 mengurangkan dimensi peta ciri secara efektif merumuskan maklumat yang paling penting dan mengurangkan kos pengiraan. Ini diikuti dengan dua lapisan konvolusi tambahan dengan penapis 64 dan 32, setiap satu diikuti dengan lapisan pengkolahan.. Lapisan-lapisan ini secara progresif mengekstrak ciri-ciri yang lebih abstrak dan lebih tinggi dari teks input.Selepas lapisan pengumpulan terakhir, lapisan rata (Flatten) digunakan untuk menukar

peta ciri berbilang dimensi pada vektor satu dimensi. Kemudian, vektor ini dihantar ke lapisan sepenuhnya bersambung dengan 256 saraf. Lapisan ini berfungsi sebagai enjin penaakulan, menggabungkan ciri-ciri yang dikesan oleh lapisan-lapisan sebelumnya dan mempelajari bagaimana mereka menyumbang kepada tugas pengelasan. Pengaktifan yang digunakan di sini ReLU.

Lapisan akhir adalah satu lagi lapisan padat dengan satu neuron output dan fungsi pengaktifan sigmoid, yang menghasilkan skor kebarangkalian antara 0 dan 1 untuk klasifikasi binari. Model ini dikompilasi dengan pengoptimum Adam, yang terkenal dengan pembelajaran adaptif, dan menggunakan binary cross entropy sebagai fungsi kehilangan, sesuai untuk tugas klasifikasi binari. Model ini dinilai menggunakan ketepatan sebagai metrik utama. Parameter tambahan dalam CNN, seperti fungsi pengaktifan (ReLU dan Sigmoid) juga membantu mengawal aliran data dan dapat meningkatkan pembelajaran.

2.6 MODEL CNN ID DENGAN PEMILIHAN CIRI



Rajah 3 Model pemilihan ciri berasaskan statistik dan klasifikasi CNN

Model CNN 1D garis dasar yang digunakan dengan pantas menyesuaikan URL latihan dengan ketepatan yang hampir sempurna tetapi menunjukkan prestasi yang lebih teruk pada data yang tertunda dimana ia menandakan pemasangan berlebihan. CNN terkenal pada input berdimensi tinggi dan jarang cenderung untuk menghafal noise melainkan dikekang. Untuk menangani perkara ini, secara teorinya menggunakan pemilihan ciri chi-kuasa dua untuk mengurangkan set token URL dan mengurangkan kerumitan model.

Pemilihan ciri ialah langkah penting dalam rangka kerja pengelasan teks. Mengalih keluar token yang menyumbang sedikit untuk membezakan pancingan data daripada URL jinak boleh mengurangkan kerumitan model dan masa latihan sambil meningkatkan ketepatan (Gaurav et al., 2024). Untuk kajian ini Chi kuasa dua telah dipilih sebagai pemilih ciri kerana ia merupakan ujian statistik yang digunakan secara meluas untuk mengukur perkaitan antara setiap token dan label kelas. Kaedah ini sangat sesuai untuk data teks yang diwakili sebagai kiraan atau kejadian binari token, dan ia berfungsi dengan cekap pada set ciri yang besar (Toğaçar, 2025). Dengan meletakkan

token mengikut skor chi kuasa duanya dimana ia menunjukkan betapa kuatnya kehadiran token berkorelasi dengan pancingan data atau label yang sah.

2.7 PENILAIAN PRESTASI

Untuk mengukur ketepatan keputusan model yang digunakan, empat jenis langkah penilaian akan digunakan iaitu positif benar, negatif benar, positif palsu, dan nilai negatif palsu. Dalam kajian ini, hasil ramalan diperiksa dengan menggunakan matrik termasuk kejituan (precision), dapatan semula (recall), ketepatan (accuracy) serta lengkung AUC-ROC. Sebagai tambahan kepada matrik ini, kajian juga mempertimbangkan pengiraan masa latihan akan direkodkan bagi membuat perbandingan model yang digunakan. Dengan menggunakan langkah penilaian tradisional ini ditakrifkan secara ringkas di bawah, di mana TP menunjukkan positif benar, TN ialah negatif benar, FP ialah positif palsu, dan FN ialah negatif palsu:

- Kejituan (Precision) : Ia memberi perkadaran positif benar kepada bilangan jumlah positif bahawa model meramalkan dan ditakrifkan seperti berikut :

$$\text{Kejituan (Precision)} = \frac{TP}{(TP + FP)}$$

- Dapatan semula (Recall) : Dapatan semula algoritma ramalan ialah bilangan ramalan URL pancingan data yang betul yang dibuat ke atas semua URL dalam set data.

$$\text{Dapatan semula (Recall)} = \frac{TP}{(TP + FN)}$$

- Ketepatan (Accuracy) : Ini mewakili bilangan kejadian data yang dikelaskan dengan betul sepanjang jumlah bilangan kejadian data dan ditakrifkan seperti berikut:

$$\text{Ketepatan (Accuracy)} = \frac{(TP + TN)}{(FP + FN + TP + TN)}$$

- Skor F1: Ia mengambil kira kejituan dan dapatan semula dan ditakrifkan seperti berikut:

$$\text{Skor F1} = 2\ \frac{(Dapatan\ semula\ \times Kejituan)}{(Dapatan\ semula + Kejituan)}$$

- Lengkung AUC-ROC: AUC - Lengkung ROC memberitahu sejauh mana model mampu membezakan antara kelas binari. Keluk ROC diplot dengan Kadar Positif Benar (TPR) terhadap Kadar Positif Palsu (FPR) di mana TPR di paksi-y dan FPR berada pada paksi-x

$$TPR = \frac{TP}{TP + FN}$$

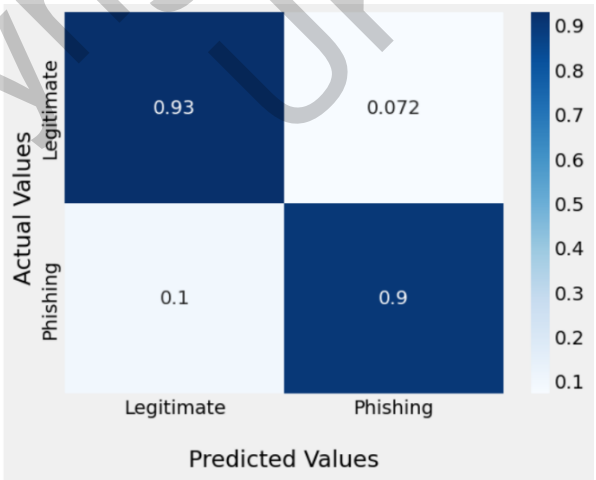
$$FPR = 1 - SEPECIFICITY = \frac{FP}{FP + TN}$$

3.0 KEPUTUSAN

Keputusan eksperimen yang dijalankan ini terdiri daripada tiga percubaan beruruta iaitu eksperimen pertama dimana prestasi CNN 1D garis dasar ditetapkan, eksperimen kedua adalah untuk menyiasat regulasasi dengan kadar keciciran yang berbeza untuk mengurangkan pemasangan berlebihan dan eksperimen yang ketiga menggunakan pemilihan ciri chi-kuasa dua untuk mengurangkan dimensi input sebelum melatih semula CNN serta mengurangkan pemasangan berlebihan.

3.1 EKSPERIMEN 1: Model Garis Dasar CNN 1D

Model Rangkaian Neural Convolutional (CNN) asas untuk pengesanan URL pancingan data dinilai dan dibincangkan. Model ini diadaptasi daripada Bakare-Opeyemi (n.d.), yang meneroka pendekatan pembelajaran mendalam untuk mengklasifikasikan URL sebagai pancingan data atau sah. CNN garis dasar berfungsi sebagai asas untuk perbandingan prestasi dalam penyelidikan ini.



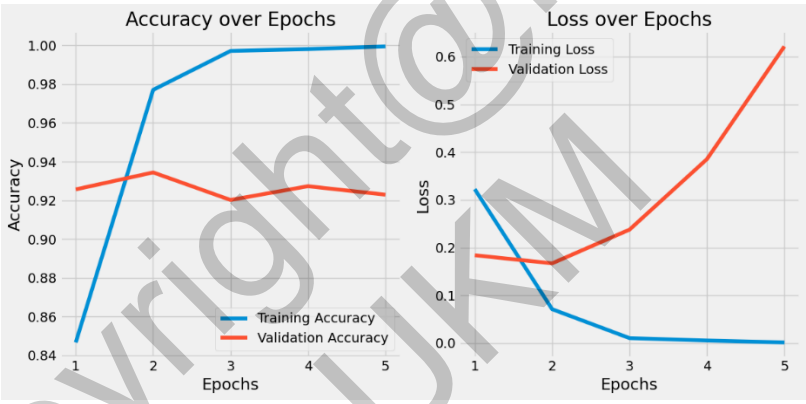
Rajah 3 Matriks kekeliruan

Jadual 1 Metrik prestasi model garis dasar

Bil Epochs	Ketepatan	Kejituan	Dapatan semula	Skor-F1	AUC	Masa/s
5	0.91	0.92	0.90	0.91	0.96	73

Berdasarkan Jadual 1, model garis dasar untuk pengesanan URL pancingan data ini telah dilatih selama 5 epoch dimana ianya mencapai prestasi ketepatan 0.91, kejituan 0.92, dapatan semula 0.90, skor F1 0.91, AUC 0.97, dan masa latihan adalah 73 saat. Keputusan ini menunjukkan bahawa model ini sangat berkesan untuk pengesanan pancingan data, menunjukkan prestasi seimbang merentas semua metrik utama. Model ini berjaya mengklasifikasikan 91% URL dengan betul menandakan tahap ketepatan yang tinggi dan sesuai digunakan sebagai garis dasar untuk kajian ini. Di samping itu, nilai kejituan yang tinggi juga menunjukkan model jarang menghasilkan positif palsu, manakala dapatan semula yang hampir setara pula membuktikan kebolehan model dalam mengesan pancingan data URL yang sebenar. Seterusnya, skor-F1 yang bernilai 0.91 yang seimbang antara kejituan dan dapatan semula menyokong kekukuhan prestasi model ini. Pada masa yang sama, nilai AUC yang hampir sempurna menunjukkan model ini sangat mampu membezakan antara pancingan data URL dan URL yang sahih.

Akan tetapi, graf kehilangan dan ketepatan model ditambah di dalam program ini untuk memberikan gambaran yang jelas tentang perilaku latihan model garis dasar ini seperti yang ditunjukkan dalam Rajah 4.



Rajah 4 Graf kehilangan dan ketepatan model

Graf ketepatan latihan meningkat daripada sekitar 85% kepada hampir 100% daripada epoch yang ketiga dan kekal stabil sehingga ke epoch yang kelima.Sebaliknya, ketepatan pengesahan memuncak pada kira-kira 92-93% selepas epoch kedua. Corak ini adalah biasa untuk model CNN yang dilatih tentang URL atau set data pancingan data, di mana model ini cepat belajar dan hampir sesuai dengan data latihan. Jurang antara latihan dan ketepatan pengesahan menandakan bahawa model itu mula melebihkan data latihan. Selain itu, graf kehilangan latihan menunjukkan bahawa kehilangan latihan menurun secara mendadak kepada hampir sifar manakala kehilangan pengesahan mula meningkat selepas epoch yang kedua selepas sedikit penurunan pada awalnya. Peningkatan dalam kehilangan pengesahan walaupun ketepatan pengesahan yang stabil adalah tanda klasik pemasangan berlebihan (overfitting), di mana model menghafal data latihan dan bukannya generalisasi pada data yang ghaib (Sudiardjo et al., 2025).

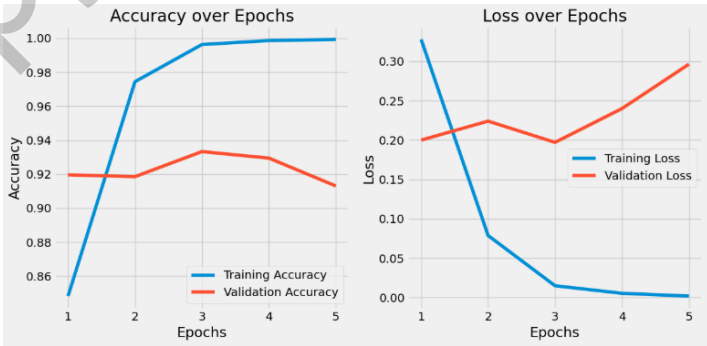
3.2 EKSPERIMEN 2: Menggunakan kadar keciciran yang berbeza

Model garis dasar tidak mempunyai lapisan tercicir dan model akan digunakan sebagai syarat kawalan untuk bandingkan dengan dari segi mengukur kesan keciciran. Untuk eksperimen ini kadar keciciran diuji dengan kadar 0.25, 0.3 dan 0.5 untuk mencegah model daripada pemasangan berlebihan dan selanjutnya menambah baik prestasi model seperti yang ditunjukkan dalam Jadual 2 (Almousa et al., 2022).

Jadual 2 Metrik prestasi dengan kadar keciciran berbeza

Kadar keciciran	Ketepatan	Kejituan	Dapatan semula	Skor-F1	AUC	Masa/s
0.25	0.91	0.89	0.94	0.91	0.97	80
0.3	0.91	0.90	0.93	0.92	0.97	60
0.5	0.90	0.93	0.86	0.89	0.97	60

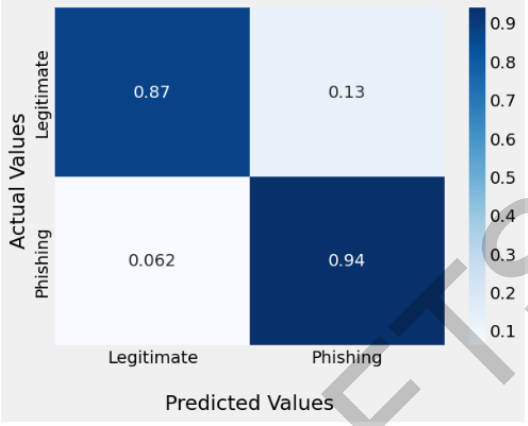
Model CNN diuji dan dinilai pada kadar keciciran 0.25, 0.3, dan 0.5. Untuk setiap tetapan, prestasi akhir pada set ujian dinilai daripada segi ketepatan, kejituan, dapatan semula, F1-skor, AUC dan masa latihan .Secara keseluruhannya, ketiga-tiga tetapan mencapai keberkesanan pengelasan yang tinggi dengan ketepatan sekitar 90-91% dan AUC = 0.97 dalam setiap kes dimana eksperimen ini menunjukkan keupayaan yang cemerlang untuk menentukan kedudukan pancingan data berbanding tapak yang sah. Walau bagaimanapun, metrik lain menunjukkan perbezaan ketara. Model dengan keciciran 0.3 memperoleh skor F1 tertinggi iaitu 0.92 dan nilai kejituan serta dapatan semula adalah seimbang, manakala keciciran 0.5 membawa kepada F1 yang lebih rendah iaitu 0.89 disebabkan kejatuhan nilai yang ketara pada nilai dapatan semula iaitu nilai turun kepada 0.86. Seterusnya, model dengan keciciran 0.25 menunjukkan dapatan semula setinggi 0.94 tetapi kejituan sedikit lebih rendah.



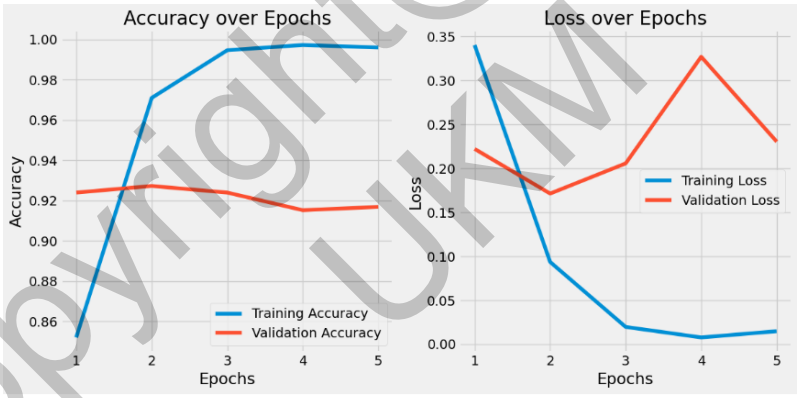
Rajah 5 Graf kehilangan dan ketepatan model dengan kadar keciciran 0.25

Dalam rajah 5 graf ketepatan menunjukkan ketepatan latihan meningkat dengan cepat, menghampiri 100%, manakala ketepatan pengesahan stabil pada nilai 0.92. Ini menunjukkan bahawa model itu sesuai dengan data latihan dan membuat generalisasi dengan baik kepada data pengesahan. Seterusnya, graf kehilangan menunjukkan kehilangan latihan menurun secara mendadak, menunjukkan pembelajaran yang berkesan. Walau bagaimanapun, kehilangan

pengesahan kekal lebih tinggi dan turun naik, menunjukkan bahawa walaupun model itu sesuai dengan data latihan dengan baik, ia mungkin masih mengalami pemasangan berlebihan pada tahap tertentu seperti yang ditunjukkan oleh kadar positif palsu yang lebih tinggi pada Rajah 6 di bawah.



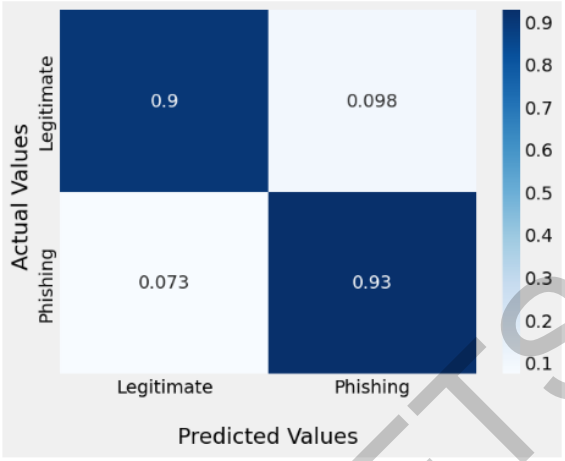
Rajah 6 Matriks kekeliruan dengan kadar keciciran 0.25



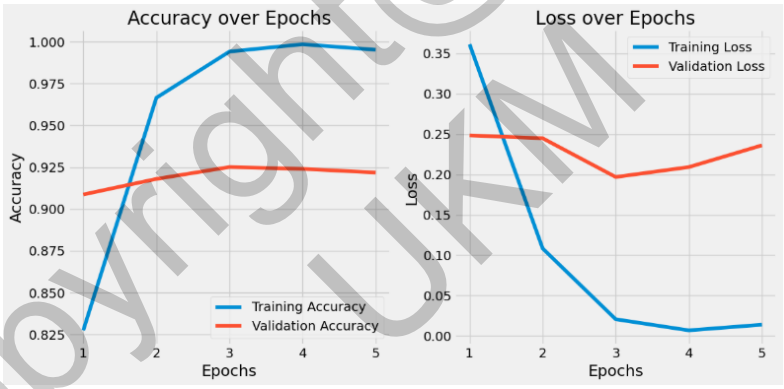
Rajah 7 Graf kehilangan dan ketepatan model dengan kadar keciciran 0.3

Dalam Rajah 7 menunjukkan graf ketepatan dan kehilangan. Pada graf ketepatan, Ketepatan latihan meningkat dengan tinggi dan hampir mencapai 100%. Ini menunjukkan kesesuaian yang baik pada data latihan. Ketepatan pengesahan menjadi stabil pada sekitar 0.92, menunjukkan bahawa model itu dapat digeneralisasikan dengan baik tanpa pemasangan berlebihan. Selain daripada itu, jika melihat dari graf kehilangan, Kehilangan latihan berkurangan dengan mendadak, menunjukkan penumpuan yang baik semasa latihan model, manakala kehilangan pengesahan turun naik sedikit antara epoch. Turun naik ini adalah perkara yang biasa apabila model menemui corak yang berbeza dalam set pengesahan, tetapi kehilangan akhir yang agak rendah menunjukkan generalisasi model yang baik. Matriks kekeliruan juga menunjukkan bahawa model mengesan pancingan data dengan baik iaitu

sebanyak 93% dan mengklasifikasikan tapak yang sah sebagai pancingan data 9.8% ditunjukkan kadar Positif Palsu di Rajah 8.



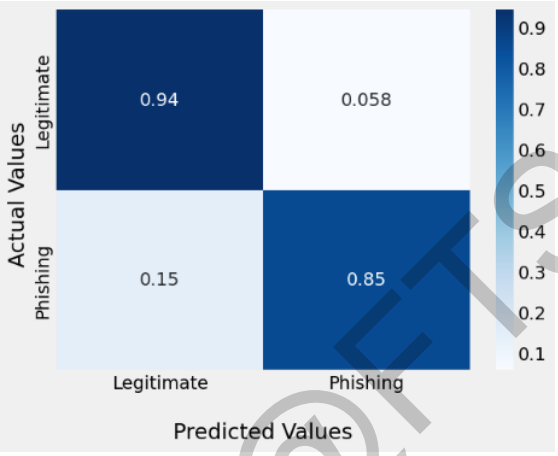
Rajah 8 Matriks kekeliruan dengan kadar keciciran 0.3



Rajah 9 Graf kehilangan dan ketepatan model dengan kadar keciciran 0.5

Pada Rajah 9 graf ketepatan menunjukkan ketepatan latihan menunjukkan peningkatan mendadak pada mulanya, menunjukkan bahawa model mempelajari data dengan cepat. Walau bagaimanapun, dataran tinggi ketepatan pengesahan pada sekitar 0.90, menunjukkan model telah mencapai had generalisasinya, dan mungkin terlalu teratur oleh kadar keciciran yang tinggi. Selain itu, pada graf kehilangan dimana kehilangan latihan berkurangan pada mulanya tetapi kekal agak rendah selepas 2 epoch. Ini adalah petanda yang baik kerana ia menunjukkan bahawa model itu sesuai dengan baik. Walau bagaimanapun, kehilangan pengesahan kekal lebih tinggi dan turun naik pada epoch yang seterusnya di mana ianya menyokong kemungkinan keciciran 0.5 mungkin telah menyebabkan beberapa kekurangan yang membawa kepada prestasi yang lemah pada data yang tidak kelihatan. Model dengan kadar keciciran 0.5

ini ia dapat mengenal pasti tapak yang sah namun ia dapat mengenal pasti lebih banyak tapak pancingan data seperti yang ditunjukkan oleh nilai negatif palsu pada Rajah 10 di bawah . Pertukaran ini menunjukkan bahawa model tidak sesuai disebabkan oleh kadar keciciran yang tinggi (0.5), yang mungkin menyebabkan model tidak dapat menangkap semua corak yang diperlukan untuk pengesanan pancingan data.



Rajah 10 Matriks kekeliruan dengan kadar keciciran 0.5

3.3 EKSPERIMEN 3: Model CNN dengan pemilihan ciri

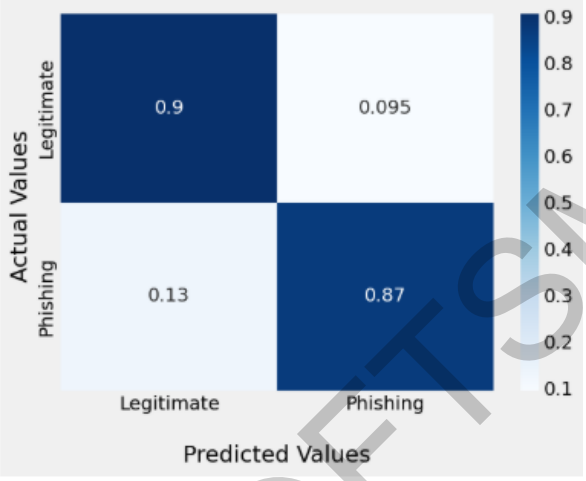
Eksperimen ini melatih epoch antara 5,10,15 dan 20 seperti yang ditunjukkan di Jadual 3 dengan menggunakan nilai keciciran 0.3 dipilih kerana kestabilan yang ditunjukkan dalam metrik prestasi dalam jadual 2.

Jadual 3 Metrik prestasi model dengan pemilihan ciri

Bil Epochs	Ketepatan	Kejituan	Dapatan semula	Skor-F1	AUC	Masa/s
5	0.88	0.92	0.84	0.87	0.96	61
10	0.88	0.91	0.84	0.87	0.95	108
15	0.89	0.90	0.86	0.88	0.96	163
20	0.89	0.9	0.87	0.88	0.95	197

Sepanjang siri latihan yang berlangsung dari 5 hingga 20 epoch, model ini secara konsisten mencapai ketepatan sekitar 88 sehingga 89%, dengan kejituan yang agak kukuh dan skor dapatan semula dengan julat kejituan 0.90 hingga 0.92, dan dapatan semula 0.84 hingga 0.87. Kejituan ini yang menunjukkan bahawa apabila model meramalkan URL adalah pancingan data, ia meramal dalam 90% betul iaitu, kadar positif palsu yang rendah. Dapatan semula sekitar 0.85 bermakna model ini menangkap kira-kira 85% daripada URL pancingan data sebenar dan kehilangan kira-kira 15% (negatif palsu). Keseimbangan ini agak baik untuk

aplikasi keselamatan, kerana ia meminimumkan penggera palsu sambil masih mengesan kebanyakan serangan. Kajian ini juga mengira Kawasan Di Bawah Lengkung ROC (AUC), yang bernilai daripada 0.95 sehingga 0.96, menunjukkan keupayaan diskriminasi keseluruhan yang sangat baik antara pancingan data dan URL yang sah.



Rajah 11 Matriks kekeliruan model dengan pemilihan ciri

Matriks kekeliruan seperti yang ditunjukkan Rajah 11 pada model ini pada epoch 5-20. Ia menunjukkan kiraan positif benar yang kuat (pancingan data dikenal pasti dengan betul) dan kiraan negatif benar (sah dikenal pasti dengan betul), dengan positif palsu dan negatif palsu yang agak kurang, sejajar dengan nilai kejituan dan dapatan semula. Selain itu , skor F1 model juga adalah tinggi sekitar 0.87–0.89, mencerminkan keseimbangan yang baik.

4.0 PERBINCANGAN

Secara keseluruhan, model CNN 1D yang dibangunkan menunjukkan prestasi tinggi dalam mengesan URL pancingan data. Model garis dasar tanpa kadar kecicir an dan tanpa pemilihan ciri mencapai ketepatan sekitar 91%, dengan kejituan 0.92 dan dapatan semula 0.90 seperti dalam Jadual 4.2. Ini bermakna model berjaya mengklasifikasikan 91% URL dengan betul serta menunjukkan keseimbangan yang baik antara kesilapan positif palsu dan negatif palsu. Nilai AUC yang menghampiri 0.97 turut memperlihatkan keupayaan pemisahan kelas yang cemerlang model hampir sempurna membezakan antara URL pancingan data dan URL sahih. Keputusan garis dasar ini konsisten dengan pencapaian tinggi .

Namun, analisis graf latihan dan pengesahan pada Rajah 4 menunjukkan tanda pemasangan berlebihan pada model garis dasar. Ketepatan latihan meningkat hampir 100% menjelang epoch ke-3, manakala ketepatan pengesahan memuncak sekitar 92 hingga 93% dan kemudian menjadi mendatar. Lebih membimbangkan, kehilangan pengesahan mula meningkat selepas epoch kedua walaupun kehilangan latihan terus menurun hampir ke sifar. Pola ini

menandakan model menghafal data latihan dan kurang menggeneralisasi corak kepada data baharu. Fenomena sebegini lazim dilaporkan dalam model pembelajaran mendalam berkapasiti tinggi. Dalam konteks kajian ini, walaupun prestasi asas tinggi, kewujudan jurang ketara antara ketepatan latihan dan pengesahan menyerlahkan keperluan untuk teknik regularisasi bagi meningkatkan kestabilan model pada data baharu.

Eksperimen 2 menyelidiki impak kadar keciciran berbeza terhadap prestasi model CNN, khususnya dalam menangani isu pemasangan berlebihan yang dikenal pasti. Keciciran sememangnya diiktiraf sebagai kaedah regularisasi berkesan untuk rangkaian neural dimana ia bertindak dengan menyahaktifkan secara rawak sebahagian neuron semasa latihan, sekali gus menghalang neuron daripada terlalu "bergantung" antara satu sama lain. Pendekatan ini memaksa model mempelajari pola yang lebih umum dan kukuh. Dalam kajian ini, tiga kadar keciciran telah diuji – 0.25, 0.3, dan 0.5 – selaras dengan amalan lazim yang menetapkan kadar keciciran sekitar 20% hingga 50%. Hasil eksperimen menunjukkan bahawa kesemuanya berjaya mengekang pemasangan berlebihan sehingga tahap tertentu, namun prestasi terbaik dicapai pada kadar keciciran 0.3.

Dengan kadar keciciran 0.3, model mencapai ketepatan 91% mempunyai nilai yang sama dengan model asas dengan skor F1 tertinggi 0.92 dalam Jadual 2. Menariknya, dapatan semula meningkat kepada 0.93 berbanding 0.90 bagi model asas, menunjukkan model dengan kadar keciciran sederhana ini lebih sensitif mengesan laman pancingan data. Kejituan pula kekal tinggi sekitar 0.90, hanya sedikit rendah berbanding model asas. Seterusnya, peningkatan dapatan semula disertai sedikit penurunan kejituan lazim berlaku apabila regularisasi berjaya membuat model lebih umum dan model berani menandakan lebih banyak kes sebagai "pancingan" . Namun dengan sedikit kompromi pada ketepatan kategori tersebut (Almousa et al., 2022). Skor F1 yang lebih tinggi membuktikan imbangan keseluruhan yang lebih baik dicapai pada kadar keciciran 0.3.

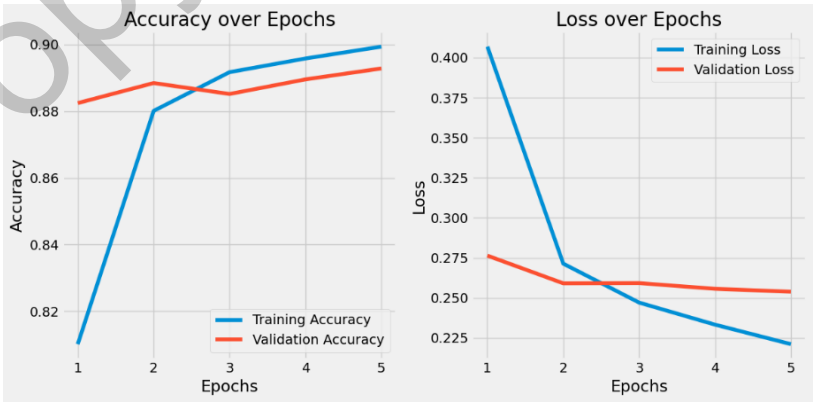
Graf prestasi untuk kadar keciciran 0.3 pada Rajah 6 memperlihatkan ketepatan pengesahan stabil 92% merentasi epoch, tanpa jurang besar dari ketepatan latihan. Malah, tiada peningkatan ketara dalam kehilangan pengesahan setelah epoch kedua yang berbeza dengan model asas yang menunjukkan pemasangan berlebihan dengan jelas. Ini menandakan regularisasi melalui kadar keciciran berjaya di mana model kurang menghafal data spesifik latihan dan lebih mampu menyesuaikan diri dengan data baharu (Almousa et al., 2022). Bagi keciciran 0.25, prestasi model hampir sama dengan model asas. Dapatan semula mencapai 0.94 dimana paling tinggi antara semua. Hal ini menunjukkan model ini sangat agresif dalam mengesan pancingan data.

Walau bagaimanapun, kejituan menurun sedikit ke 0.89 berbanding 0.92 dengan model garis dasar, bermakna lebih banyak laman sah diklasifikasikan sebagai pancingan data (positif palsu meningkat). Corak ini mencadangkan bahawa kadar keciciran 0.25 mungkin masih belum mencukupi untuk mengekang kecenderungan model menghafal beberapa corak spesifik pada data latihan. Model mungkin mengalami sedikit pemasangan berlebihan pada ciri-ciri tertentu

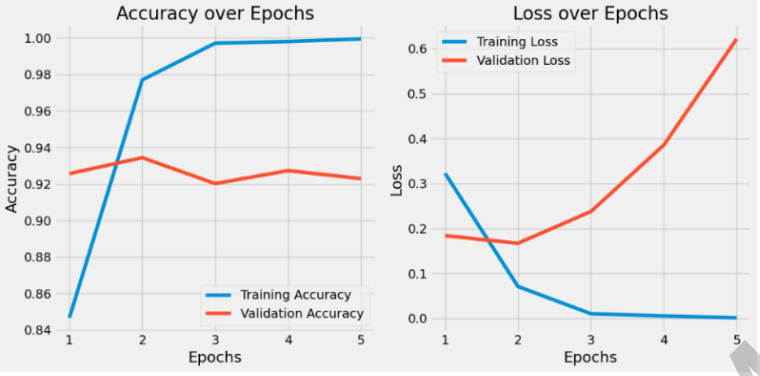
yang kerap muncul dalam pancingan data, sehingga menandakan beberapa URL sah sebagai mencurigakan. Namun begitu, perbezaannya tidak besar, dan AUC kekal bernilai 0.97, menunjukkan model 0.25 masih berupaya membezakan kelas dengan baik.

Dalam kes keciciran 0.5, kita dapat melihat kesan regularisasi berlebihan. Model dengan 50% neuron digugurkan semasa latihan menunjukkan ketepatan sedikit lebih rendah iaitu 90% dan skor F1 terendah iaitu 0.89 antara ketiga-tiga tetapan kadar keciciran. Kejituan model 0.5 sebenarnya paling tinggi 0.93, bermakna ia jarang sekali memberi amaran palsu di mana ia bermaksud indikator bahawa model ini sangat konservatif. Akan tetapi, dapatan semula merosot ke 0.86, menandakan model terlepas lebih banyak laman pancingan data yang sebenar berbanding model lain. Ini konsisten dengan pemahaman bahawa kadar keciciran yang terlalu tinggi boleh menyebabkan model terkurang latih (underfitting) iaitu model gagal mempelajari sepenuhnya corak penting kerana terlalu banyak neuron tidak aktif pada setiap iterasi latihan. Akibatnya, model 0.5 cenderung kurang sensitif terhadap variasi serangan pancingan data, walaupun ia lebih spesifik. Dalam konteks keselamatan siber, tetapan begini mungkin kurang ideal kerana negatif palsu (ancaman terlepas) adalah lebih berbahaya berbanding negatif benar.

Penemuan di atas menunjukkan keperluan imbangan regularisasi yang optimum. Kadar keciciran yang sederhana sekitar 20 sehingga 30% lazimnya memberikan keseimbangan terbaik antara bias dan varians model (Almousa et al., 2022). Secara keseluruhannya, penambahan lapisan kadar keciciran dalam model CNN garis dasar ini membantu menstabilkan pembelajaran dan mengurangkan jurang prestasi antara data latihan dan pengesahan. Model dengan keciciran 0.3 khususnya menunjukkan prestasi unggul dan akan dijadikan rujukan untuk penambahbaikan selanjutnya.



Rajah 12 Graf kehilangan dan ketepatan dengan pemilihan ciri dengan 5 epoch



Rajah 13 Graf kehilangan dan ketepatan model garis dasar dengan 5 epoch

Seterusnya, penggunaan pemilihan ciri dengan kadar keciciran 0.3 nyata mendatangkan beberapa kesan positif. Pertama, model menjadi lebih ringkas dan pantas dilatih tanpa pemasangan berlebihan. Jika melihat pada Rajah 12 dan Rajah 13, perbandingan kehilangan pengesahan pada graf kehilangan model garis dasar lebih tinggi berbanding graf kehilangan model CN 1D dengan pemilihan ciri. Hal ini menandakan bahawa eksperimen CNN 1D dengan pemilihan ciri dapat mengelakkan model daripada pemasangan berlebihan. Dengan skop ciri yang lebih kecil, kos pengiraan untuk lapisan pembenaman dan konvolusi berkurang. Model juga berpotensi kurang terdedah kepada noise daripada ciri-ciri yang tidak relevan. Dengan membuang token-token yang jarang muncul atau tidak berkait rapat dengan pancingan data, kita mengurangkan risiko model "keliru" atau terlalu memfokus pada corak kebetulan.

Model ciri terpilih mengalami sedikit penurunan dapatan semula sekitar 0.84–0.87 berbanding 0.90 model garis dasar. Ini bererti terdapat beberapa URL pancingan data yang dahulunya dikesan oleh model penuh, kini terlepas. Kemungkinan besar, token atau pola unik dalam URL pancingan data tersebut tidak termasuk dalam 1000 ciri teratas, menyebabkan model gagal mengenal pasti ciri tersebut. Dengan kata lain, pemilihan ciri yang terlalu agresif boleh mengabaikan sebahagian petunjuk serangan yang jarang berlaku tetapi penting. Fenomena "pulangan semakin berkurang" juga diakui dalam literatur dimana menambah ciri di luar lingkungan tertentu memberikan manfaat marginal, namun membuang terlalu banyak ciri boleh mula menjejaskan prestasi (Tang, 2024).

Pemilihan ciri Chi-kuasa dua memang efektif meningkatkan kecekapan klasifikasi teks, tetapi jumlah ciri optimum perlu ditentukan secara empirikal kerana terlalu sedikit ciri boleh mengurangkan kandungan maklumat klasifikasi. Oleh itu, pemilihan 1000 ciri dalam kajian ini adalah langkah berhati-hati kerana jumlah ini masih cukup besar untuk merangkumi pelbagai token yang relevan (Hokijuliandy et al., 2023). Keputusan menunjukkan pendekatan ini berhasil mengekalkan prestasi hampir setara model penuh, menandakan 1000 ciri mungkin titik imbangan yang sesuai bagi set data ini.

Dari sudut perbandingan, model dengan pemilihan ciri menghasilkan kejituan tinggi iaitu 0.88 hingga 0.89 yang konsisten merentasi epoch. Implikasinya adalah model jarang memberi amaran palsu meskipun dengan ciri yang terhad . Hal ini menunujukkan satu petunjuk

positif bahawa ciri-ciri paling signifikan untuk mengenal pasti laman sah telah dikekalkan. Nilai AUC iaitu 0.95 juga menunjukkan model masih handal membezakan kelas, hanya sedikit di belakang model penuh yang AUC-nya 0.96–0.97. Kekuatan ini penting jika sistem akan digunakan secara langsung.

5.0 KESIMPULAN

Kesimpulanya, kajian ini telah menunjukkan keberkesanan CNN 1D untuk pengesanan URL pancingan data dan meneroka cara keciciran dan pemilihan ciri boleh mengurangkan permasangan berlebihan tanpa ketepatan yang jauh berbeza. CNN garis dasar mencapai ketepatan yang tinggi dalam membezakan pancingan data daripada URL yang sah tetapi menampakkan ciri model mengalami pemasangan berlebihan. Pengaturan keciciran ini dapat membantu model mengelakkan pemasangan berlebihan dan pemilihan ciri Chi-kuasa dua dapat mengekalkan ketepatan dan kecekapannya.

.

6.0 BATASAN KAJIAN

Kajian ini mempunyai beberapa kekangan yang perlu diambil kira. Pertama, saiz dan kebolehwakilan dataset adalah terhad kerana hanya menggunakan set data yang kecil atau seimbang secara buatan, yang mungkin tidak mencerminkan kepelbagaian URL sebenar di dunia nyata dimana memerlurkan kajian yang lebih lanjut menggunakan set data yang lebih besar dan pelbagai. Kedua, kaedah pencegahan pemasangan berlebihan hanya tertumpu kepada penggunaan kadar keciciran (dropout), sedangkan teknik lain seperti regularisasi L1/L2, early stopping, dan data augmentation boleh diterokai dalam kajian akan datang untuk meningkatkan kebolehan model dalam membuat generalisasi. Selain itu, model asas dalam kajian ini hanya dilatih selama lima epoch, yang mungkin tidak mencukupi untuk mencapai konvergensi penuh. Oleh itu, kajian masa hadapan boleh meningkatkan bilangan epoch latihan bagi memperoleh prestasi yang lebih stabil dan mantap.

7.0 PENGHARGAAN

Penulis kajian ini ingin ucapkan setinggi-tinggi penghargaan dan jutaan terima kasih kepada Ts. Dr. Wan Fariza Paizi@Fauzi, penyelia penulis kajian ini yang telah memberi tunjuk ajar serta bimbingan untuk menyiapkan projek ini dengan jayanya.

Penulis kajian ini juga ingin mengucapkan terima kasih kepada semua pihak yang membantu secara langsung mahupun tidak langsung dalam menyempurnakan projek ini. Segala bantuan yang telah dihulurkan amatlah dihargai kerana tanpa bantuan mereka, projek ini tidak dapat dilaksanakan dengan baik. Semoga tuhan merahmati dan memberikan balasan yang terbaik.

8.0 RUJUKAN

Albishri, A. A., & Dessouky, M. M. (2024). A comparative analysis of machine learning techniques for URL phishing detection. *Engineering Technology & Applied Science Research*, 14(6), 18495–18501. https://doi.org/10.48084/etasr.8920

Almousa, M., Zhang, T., Sarrafzadeh, A., & Anwar, M. (2022). Phishing website detection: How effective are deep learning-based models and hyperparameter optimization? *Security and Privacy*, 5(6). https://doi.org/10.1002/spy2.256

Bakare-Opeyemi. (n.d.). GitHub - Bakare-Opeyemi/Phishing-URL-Detection: In this repository, I have used three unique deep learning architectures to develop models for the detection of phishing URLs. GitHub. https://github.com/Bakare-Opeyemi/Phishing-URL-Detection

Butnaru, A., Mylonas, A., & Pitropakis, N. (2021). Towards lightweight URL-based phishing detection. *Future Internet*, 13(6), Article 154. https://doi.org/10.3390/fi13060154

Ejaz, A., Mian, A. N., & Manzoor, S. (2023). Life-long phishing attack detection using continual learning. *Scientific Reports*, 13(1). https://doi.org/10.1038/s41598-023-37552-9

Gaurav, A., Chui, K. T., Arya, V., Attar, R. W., Bansal, S., Alhomoud, A., & Psannis, K. (2024). Optimized AI-driven semantic web approach for enhancing phishing detection in E-Commerce platforms. *International Journal on Semantic Web and Information Systems*, 20(1), 1–13. https://doi.org/10.4018/ijswis.359767

Haq, Q. E. U., Faheem, M. H., & Ahmad, I. (2024). Detecting phishing URLs based on a deep learning approach to prevent cyber-attacks. *Applied Sciences*, 14(22), 10086. https://doi.org/10.3390/app142210086

Hokijuliandy, E., Napitupulu, H., & Firdaniza, N. (2023). Application of SVM and chi-square feature selection for sentiment analysis of Indonesia's National Health Insurance mobile application. *Mathematics*, 11(17), 3765. https://doi.org/10.3390/math11173765

Opara, C., Chen, Y., & Wei, B. (2023). Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics.

S, A., & M. S., V. P. (2024). Enhancing web security: An efficient URL phishing classifier based on deep learning. *Grenze International Journal of Engineering & Technology (GIJET)*, 10(1, Part 1), 337–345.

Subasi, A., & Kremic, E. (2020). Comparison of Adaboost with MultiBoosting for Phishing Website Detection. *Procedia Computer Science*, *168*, 272–278. https://doi.org/10.1016/j.procs.2020.02.251

Sudiardjo, K. N., Alam, I. N., Wijaya, W., & Wulandhari, L. A. (2025). Diagnostic uncertainty in pneumonia detection using CNN MobileNetV2 and CNN from scratch. *arXiv.org*. https://arxiv.org/abs/2505.02396

Toğaçar, M. (2025). Integration of mobile deep networks and machine learning methods for flood risk classification: 2D grayscale transformation of data, feature intersection. *Acta Geophysica*. https://doi.org/10.1007/s11600-025-01624

Yerima, S. Y., Alzaylaee, M. K., Cyber Technology Institute, Al-Qunfudah College of Computing, Faculty of Computing, Engineering and Media, De Montfort University, & Umm Al-Qura University. (2020). High accuracy phishing detection based on convolutional neural networks. *Third International Conference on Computer Applications & Information Security (ICCAIS 2020), 19-21 March, 2020, Riyadh, Saudi Arabia*.

Nur Syazwana binti Mohd Yunan (A193884)
Ts. Dr. Wan Fariza Paizi@Fauzi
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia