PENGESANAN BULI SIBER BERASASKAN IMEJ DI MEDIA SOSIAL MENGGUNAKAN EFFICIENTNETB3 DAN PEMBELAJARAN META-ENSEMBLE

[1]Kavin Arasan A/L Mudiarasan, [1]Wandeep Kaur A/P Ratan Singh

[1]Fakulti Teknologi & Sains Maklumat
43600 Universiti Kebangsaan Malaysia

Abstrak

Buli siber telah menjadi satu ancaman yang semakin membimbangkan dalam era digital, khususnya dalam platform media sosial berasaskan imej di mana kandungan visual yang berunsur negatif boleh memberikan kesan psikologi yang serius. Kajian ini menangani masalah penting dalam mengesan imej yang mengandungi unsur buli siber dengan meneroka teknik pembelajaran mendalam (*deep learning*) dan pembelajaran mesin (*machine learning*) bagi mengklasifikasikan kandungan visual dengan lebih tepat. Objektif kajian ini adalah untuk membangunkan satu sistem pengelasan imej buli siber yang cekap dengan menggabungkan analisis visual melalui pengekstrakan ciri dan algoritma klasifikasi. Kajian ini dijalankan menggunakan set data imej yang telah dikategorikan sebagai "buli siber" dan "bukan buli siber". Proses penyelidikan melibatkan prapemprosesan imej menggunakan CLAHE (*Contrast Limited Adaptive Histogram Equalization*), penskalaan ciri dengan MinMaxScaler, dan pengurangan dimensi menggunakan PCA (Principal Component Analysis) bagi meningkatkan prestasi model. EfficientNetB3 digunakan sebagai rangkaian neural konvolusi pra-latih (*pre-trained*) untuk mengekstrak corak visual yang bermakna, yang kemudiannya diklasifikasikan menggunakan meta-pengelasan (*meta-classifier*) gabungan SVM-XGBoost. Hasil eksperimen menunjukkan bahawa model yang dipertingkat ini mencapai nilai *accuracy* sebanyak 99.17% yang jauh mengatasi model asas (*EfficientNetB0-SVM),* serta menunjukkan nilai precision, recall, dan F1-score yang tinggi bagi kedua-dua kelas. Eksperimen pengablasian (*ablation experiment*) turut membuktikan kepentingan setiap ciri dan proses prapemprosesan dalam meningkatkan ketepatan dan keupayaan generalisasi model. Kajian ini menyumbang kepada bidang pengkomputeran dengan memperkenalkan seni bina hibrid yang berskala dan boleh dijelaskan untuk pengesanan automatik buli siber berasaskan imej, yang boleh diintegrasikan ke dalam aplikasi dunia sebenar seperti sistem pemantauan kandungan di platform media sosial. Hasil kajian ini turut memberi implikasi

terhadap dasar platform dengan membantu mengenal pasti kandungan berbahaya secara automatik, memupuk interaksi digital yang lebih selamat, dan menyokong intervensi awal dalam menangani buli siber.

Kata Kunci: pembelajaran mendalam, pembelajaran mesin, prapemprosesan, penskalaan ciri

Abstract

Cyberbullying has become an alarming and pervasive threat in the digital age, especially on image-based social media platforms where harmful visual content can inflict significant psychological damage. This study addresses the critical problem of detecting cyberbullying in images by investigating advanced deep learning and machine learning techniques to classify visual content accurately. The aim of the study is to develop an effective cyberbullying image classification system that incorporates visual analysis through feature extraction and advanced classification algorithms. The research explores the use of an ensemble learning model integrating EfficientNetB3 for feature extraction with an SVM-XGBoost meta-classifier to enhance classification accuracy and robustness. Conducted using a curated dataset comprising images labeled as "cyberbullying" and "non-cyberbullying," the methodology involved image preprocessing using CLAHE (Contrast Limited Adaptive Histogram Equalization), feature scaling via MinMaxScaler, and dimensionality reduction using PCA to optimize model performance. EfficientNetB3 served as a pre-trained convolutional neural network to extract meaningful visual patterns, which were then classified using the stacked meta-classifier. Experimental results demonstrated that the enhanced model achieved a high test accuracy of up to 99.17%, significantly outperforming the base model EfficientNetB0-SVM and exhibiting strong precision, recall, and F1-scores across both classes. Ablation experiments further validated the importance of each feature and preprocessing enhancement in improving the model's accuracy and generalization. This research contributes to the field by presenting a scalable and explainable hybrid architecture for automated image-based cyberbullying detection, which can be integrated into real-world applications, such as content moderation systems in social media platforms. The findings have potential implications for platform policy enforcement, as automated systems can assist in flagging harmful content, promoting safer digital interactions, and supporting early intervention strategies. The model thus offers a valuable tool for mitigating online harm through intelligent image classification.

1.0 PENGENALAN

Buli siber ialah satu bentuk gangguan yang berlaku secara dalam talian melalui platform digital, termasuk media sosial, aplikasi pemesejan, forum, dan permainan dalam talian. Ia melibatkan penggunaan teknologi untuk sengaja mengugut, atau memalukan seseorang individu. Berbanding dengan perbuatan buli secara fizikal, buli siber boleh berlaku pada bila-bila masa dan tersebar secara meluas. Malahan. kesan buli siber juga boleh menjadi teruk dan berpanjangan. Penyelidikan terkini menekankan bahawa mangsa buli siber sering mengalami kesan negatif terhadap kesihatan mental, termasuk peningkatan kebimbangan, kemurungan, dan juga pemikiran untuk membunuh diri (Claudio Longobardi dan Tomas Jungert, 2023). Kesan buruk terhadap akademik dan sosial juga menjadi salah satu impak buli siber kerana mangsa mungkin menjauhkan diri daripada sekolah atau aktiviti sosial akibat ketakutan atau rasa malu (Ortega-Barón et al., 2020).

Menurut kajian tahun 2020 oleh Cyberbullying Research Center, sebanyak 34% remaja di Amerika Syarikat melaporkan bahawa mereka pernah mengalami buli siber pada suatu ketika dalam hidup mereka (Patchin & Hinduja, 2020). Indeks Kesejahteraan Belia Global (Global Youth Wellbeing Index) menunjukkan bahawa buli siber ialah isu global, dengan sekitar 30% belia di beberapa negara melaporkan gangguan dalam talian (Youth Wellbeing Index, 2021). Statistik ini menggambarkan kesan buli siber yang meluas dan berbahaya terhadap golongan anak muda di seluruh dunia. Justeru, gejala ini menekankan keperluan untuk meningkatkan kesedaran dan strategi intervensi. Media sosial memainkan peranan penting dalam peningkatan buli siber kerana menyediakan platform untuk mesej yang boleh mengganggu individu dengan menyampaikan khabar angin atau imej yang menyakitkan hati. Media sosial juga memeberi kesan yang cepat dan besar kepada khalayak yang besar (Maryam Ammar Lutf Al-Anesi, 2022).

Berbanding dengan interaksi bersemuka, media sosial membolehkan pembuli mengekalkan rahsia identiti mereka dan mendorong mereka untuk bertindak dengan lebih agresif kerana akauntabiliti yang berkurangan (Hinduja & Patchin, 2020). Identiti tanpa nama digabungkan dengan akses berterusan yang disediakan oleh media sosial, menyebabkan buli siber boleh berlaku sepanjang masa dan menjadikan mangsa sukar untuk melarikan diri daripada gangguan. Laporan tahun 2021 oleh Cyberbullying Research Center mendapati bahawa peratusan yang signifikan daripada remaja di Amerika Syarikat telah mengalami beberapa bentuk buli siber, dengan platform media sosial seperti Instagram, Facebook, dan Snapchat sering disebut sebagai laman utama bagi insiden ini (Patchin & Hinduja, 2021). Tambahan pula, kajian menunjukkan bahawa sifat umum media sosial

melanjutkan kesan emosi yang ditimbulkan oleh buli siber, kerana komen atau imej yang menyakitkan hati boleh menyebabkan rasa malu yang meluas dan pengasingan sosial bagi mangsa tersebut (Pabian & Vandebosch, 2021). Oleh itu, hasil dapatan ini menekankan bagaimana media sosial boleh menyumbang kepada kekerapan dan intensiti buli siber, sekaligus meningkatkan beban psikologi terhadap pengguna muda.

2.0 KAJIAN LITERATUR

Kajian Model Pembelajaran mesin dalam pengesanan imej buli siber

Analisis kajian-kajian terdahulu melibatkan penemuan dan pemilihan artikel penyelidikan di mana kertas penyelidikan yang paling banyak dirujuk berkaitan buli siber dan pembelajaran mesin telah dikumpulkan.

(Ammar Almomani dan Khalid Nahar , 2024) mencadangkan model pembelajaran mendalam yang menggunakan seni bina Convolutional Neural Network (CNN), khususnya VGG16 dan ResNet50 untuk pengekstrakan ciri. Ciri-ciri yang diekstrak kemudiannya dimasukkan ke dalam pengklasifikasi pembelajaran mesin seperti SVM dan Logistic Regression untuk pengkategorian. Model tersebut mencapai ketepatan sebanyak 67% dalam kandungan pembulian eksplisit. Namun, model ini menghadapi kesukaran dalam memahami konteks di mana buli siber sering melibatkan simbol budaya atau masyarakat, isyarat, atau imej di mana maksudnya boleh menjadi berbeza mengikut latar belakang penonton. Sebagai contoh, imej yang mengandungi isyarat halus seperti ekspresi mengejek mungkin kelihatan tidak berbahaya tanpa memahami kepentingan budaya atau sosialnya. Selain itu, meme atau suntingan visual mungkin menggunakan sindiran atau humor untuk menyembunyikan niat jahat. Ini menjelaskan ketepatan rendah model tersebut.

Seterusnya, (Tofayet Sultan dan Nusrat Jahan , 2023) mencadangkan model untuk mengesan buli siber berasaskan imej dengan menggunakan Optical Character Recognition (OCR) untuk mengekstrak teks daripada imej dan algoritma pembelajaran mesin berasaskan NLP seperti Logistic Regression (LR) dan SVM untuk pengklasifikasian. Hasil menunjukkan model tersebut mencapai ketepatan 96%. Malangnya, sistem ini hanya memproses kandungan teks yang diekstrak daripada imej dan tidak mengambil kira konteks visual, seperti isyarat, simbol, atau petunjuk visual yang terdapat dalam imej. Sebagai contoh, isyarat seperti menunjukkan jari tengah atau ibu jari ke bawah tergolong dalam buli siber, tetapi sistem tidak dapat mengenal pasti dan mengesannya.

Beberapa alat sedia ada seperti OpenPose dan Google Cloud Vision API digunakan oleh (Nishant Vishwamitra dan Hongxin Hu , 2021) untuk pengekstrakan ciri bagi mengesan postur badan yang menghina dan pengenalan isyarat tangan. Penulis kemudian menggunakan CNN multimodal-(Multi-layer perceptron (MLP)) yang mencapai ketepatan 93.36%. Namun, model seperti ini tidak mampu mengenal pasti buli siber dalam imej yang mengandungi unsur keterlanjuran atau keganasan dengan berkesan. Ini boleh menjejaskan prestasi model ketika mengesan hantaran buli siber di pelbagai platform media sosial.

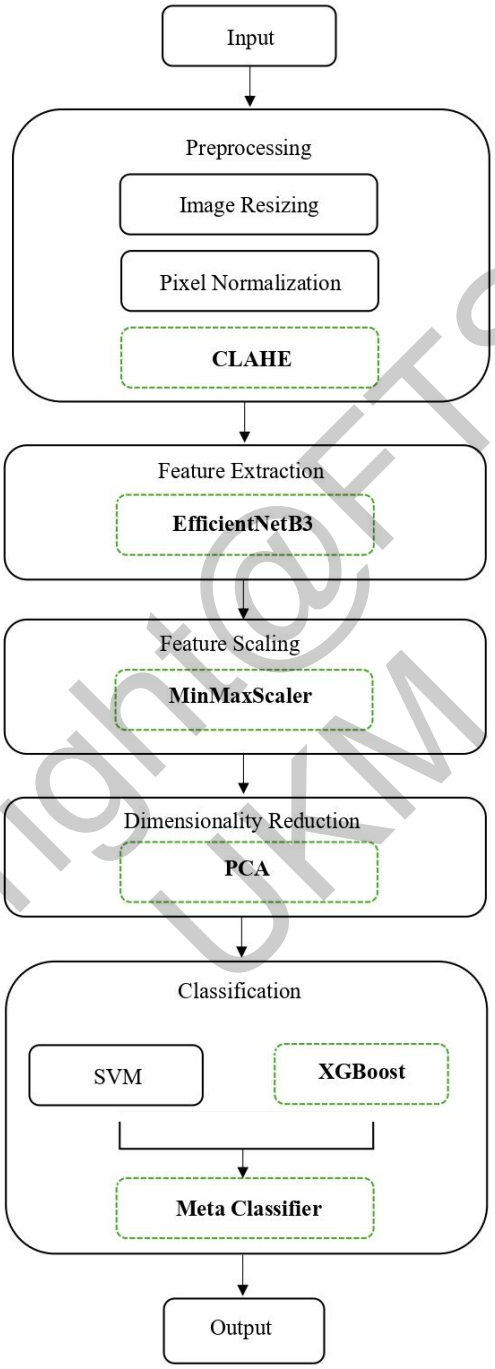
Selain itu, (Pradeep Kumar Roy dan Fenish Umeshbhai Mali , 2022) menggunakan 2D CNN dan model pembelajaran pemindahan seperti VGG16 dan InceptionV3 untuk mengekstrak ciri kompleks daripada imej. Hasilnya, InceptionV3 mengatasi VGG16 dengan ketepatan 89%. Kajian ini menunjukkan bahawa model pembelajaran pemindahan seperti InceptionV3 sangat sesuai untuk pengesanan buli siber berasaskan imej. Walau bagaimanapun, model ini hanya mengesan tanda buli siber yang eksplisit, seperti imej atau teks yang jelas menghina dalam imej.

Begitu juga, (Md. Tarek Hasan dan Md. Al Emran Hossain , 2023) menggunakan model pembelajaran mendalam (CNN) untuk mengenal pasti elemen ofensif atau berbahaya. Model pembelajaran mendalam (DL) menunjukkan peningkatan ketepatan berbanding kaedah pembelajaran mesin tradisional dalam mengenal pasti kandungan visual yang menghina. Namun, kedua-dua model menghadapi kesukaran untuk mengesan petunjuk tersirat, seperti bahasa badan yang halus, isyarat bergantung konteks, atau elemen simbolik yang menyampaikan niat jahat. Sebagai contoh, emoji seperti mata menjeling atau senyuman sinis pada imej tidak dikesan sebagai buli siber, yang menjejaskan ketepatan keseluruhan model.

3.0 METODOLOGI

Kajian ini membina model EfficientNetB3-SVM-XGBoost yang digunakan untuk mengesan imej buli siber yang diambil daripada media sosial seperti Instagram, Snapchat dan TikTok. Model ini berfungsi dengan mengekstrak ciri daripada imej set data menggunakan model EfficientNetB3 terlatih terdahulu (pre-trained). Seterusnya, ciri-ciri yang diekstrak akan diklasifikasikan sama ada imej tersebut adalah buli siber atau tidak melalui kaedah Stacking Ensemble yang menggabungkan pengelas SVM dan XGBoost, dengan pengelas meta Ridge Classifier.

3.1 Reka bentuk Model



Berdasarkan Rajah di atas , model yang dicadangkan masih mengekalkan struktur model asas iaitu hibrid antara CNN dan pengelas pembelajaran mesin. Namun, ciri-ciri yang ditandakan dalam kotak hijau merupakan penambahbaikan yang dibuat untuk meningkatkan kecekapan dan prestasi model. Model ini dimulakan dengan input, iaitu imej mentah dimuat naik daripada set data yang terdiri daripada dua kategori di mana imej berkaitan buli dilabelkan sebagai (Ya) dan bukan buli dilabelkan sebagai (Tidak).

Seterusnya, peringkat prapemprosesan dilaksanakan untuk meringkaskan imej bagi meningkatkan ketepatan pengekstrakan ciri. Proses ini melibatkan penyesuaian saiz imej kepada 300×300 piksel untuk memastikan keseragaman dan kesesuaian dengan model pembelajaran mendalam. Normalisasi piksel dilakukan bagi menyesuaikan julat nilai piksel dengan keperluan EfficientNetB3. Tambahan pula, teknik Contrast Limited Adaptive Histogram Equalization (CLAHE) digunakan untuk meningkatkan kontras imej agar ciri-ciri utama lebih mudah dikenal pasti. Setelah selesai prapemprosesan, imej dihantar ke peringkat pengekstrakan ciri, di mana EfficientNetB3 sebagai CNN yang berkuasa akan mengekstrak representasi ciri yang mendalam daripada lapisan konvolusi tertinggi. Ciri-ciri yang telah diekstrak kemudian dinormalisasi menggunakan MinMaxScaler bagi menstabilkan julat nilai antara 0 hingga 1 untuk mengelakkan dominasi oleh ciri bermagnitud tinggi. Selepas itu, pengurangan dimensi dilakukan menggunakan Principal Component Analysis (PCA), yang mengekalkan 98% variasi data sambil mengurangkan kerumitan pengiraan dan membuang maklumat berlebihan.

Dengan ciri yang telah dioptimumkan, model memasuki fasa pengelasan, di mana dua pengelas pembelajaran mesin, Support Vector Machine (SVM) dan XGBoost, dilatih secara berasingan untuk mengklasifikasikan imej kepada kategori buli atau bukan buli. Keputusan klasifikasi daripada kedua-dua pengelas ini akan dimasukkan ke dalam pengelas meta (Stacking Ensemble) yang menggabungkan kekuatan kedua-dua model untuk meningkatkan ketepatan dan daya tahan. Sebagai perbandingan, model asas hanya menggunakan SVM sebagai satu-satunya pengelas. Akhirnya, model ini akan menghasilkan output iaitu klasifikasi sama ada "Ya" (buli siber dikesan) atau "Tidak" (tiada buli siber), sekali gus menyediakan sistem pengesanan buli siber yang berkesan dan berstruktur. Prestasi model akan diuji menggunakan pelbagai metrik klasifikasi seperti precision, recall, accuracy, specificity, F1-score, dan juga melalui kaedah K-Fold cross-validation.

Kerangka yang dicadangkan ini merupakan versi yang dipertingkatkan daripada model yang dicadangkan oleh (Mahmoud Elmezain dan Amer Malki , 2022). Mereka menyatakan bahawa model pra-latih seperti EfficientNet dapat mengekstrak ciri visual peringkat tinggi yang merangkumi maklumat paling relevan daripada imej, walaupun dalam set data yang mempunyai konteks khusus seperti buli siber.

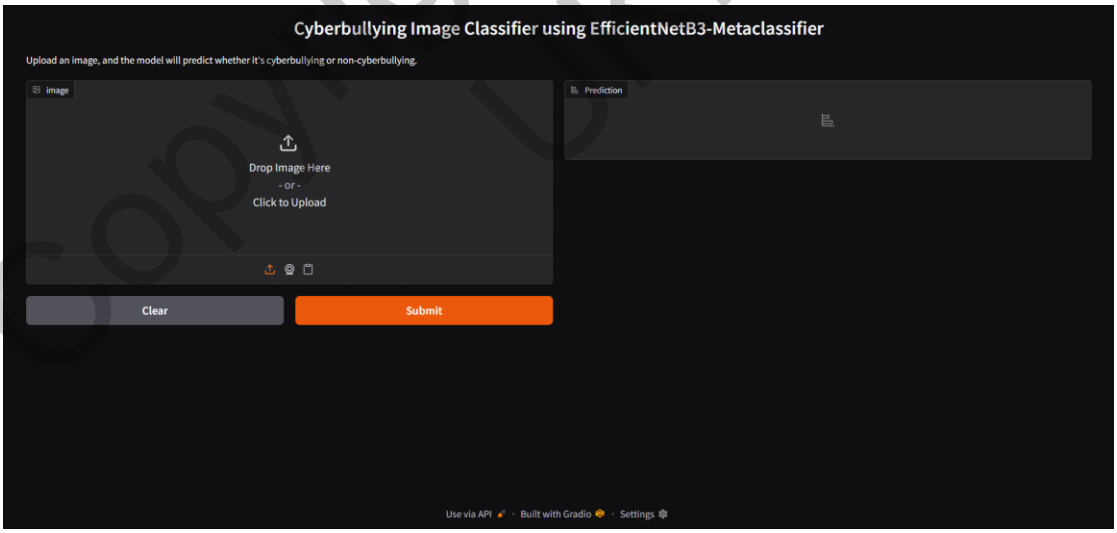
Model ini dibandingkan dengan model pengesanan buli siber berasaskan imej oleh (Mahmoud Elmezain dan Amer Malki , 2022) yang menggunakan EfficientNetB0 dan SVM. Model hibrid ini sering dijadikan garis dasar dalam tugasan klasifikasi berasaskan imej kerana keberkesanan dan tahap kecekapannya. Oleh itu, perbandingan ini menyerlahkan kecekapan model yang dipertingkat, kerana ia mampu mencapai ketepatan yang lebih baik dengan bilangan parameter yang lebih rendah.

subtopik di dalam aplikasi, serta soalan yang perlu diterapkan ke dalam aplikasi.

Hasil temubual telah mengubah sedikit beberapa keperluan pengguna. Pensyarah mencadangkan lebih banyak soalan terarah bagi meningkatkan kefahaman pelajar. Selain itu, pengasingan kuiz mengikut subtopik perlu dibangunkan bagi membantu pelajar yang lemah. Menurut pensyarah, lebih baik jika bilangan soalan diperbanyakkan dalam aplikasi yang dibangun. Aplikasi akan menerapkan fungsi kuiz secara multi-pengguna bagi mewujudkan persaingan secara sihat dan menimbulkan motivasi terhadap pelajar.

Selain daripada mendapatkan keperluan daripada pemegang taruh, analisis aplikasi yang tersedia juga dilakukan bagi memperoleh keperluan tambahan. Aplikasi yang dianalisis adalah *Dat Thin Pone High School Biology AR Learning* yang mana aktiviti di dalam aplikasi tersebut mengandungi beberapa kategori iaitu realiti terimbuh, multimedia, ujian, topik, grafik dan simulasi makmal. Keperluan aplikasi diperoleh dengan mengadaptasi beberapa ciri yang terdapat pada aplikasi yang dinyatakan seperti multimedia, ujian dan bab. Konsep pembelajaran melalui unsur grafik yang menarik berserta persekitaran realiti terimbuh dan konsep gamifikasi akan diterapkan di dalam aplikasi untuk meningkatkan kualiti pembelajaran.

3.2 Reka Bentuk Antara Muka Model



Rajah 2 reka bentuk antara muka *Gradio*

Antara muka Gradio memainkan peranan penting dalam model ini dengan menyediakan antara muka hadapan yang mesra pengguna dan interaktif untuk pengelasan imej buli siber secara masa nyata. Selain itu, Gradio membolehkan pengguna memuat naik imej terus melalui pelayar web dan menerima keputusan sama ada imej tersebut tergolong dalam kategori "Buli siber" atau "Bukan Buli siber" secara serta-merta. Ini secara signifikan memudahkan proses pengujian model dan meningkatkan kebolehaksesan untuk aplikasi dunia sebenar. Dalam projek ini, antara muka Gradio diintegrasikan dengan komponen model latar belakang, termasuk EfficientNetB3 untuk pengekstrakan ciri, MinMaxScaler dan PCA untuk pemprosesan ciri, serta model himpunan bertindan yang merangkumi SVM, XGBoost, dan Ridge Classifier.

Apabila imej dimuat naik, ia akan melalui saluran prapemprosesan yang sama (termasuk CLAHE dan pelarasan saiz), dan outputnya dihantar melalui model yang telah dilatih untuk menghasilkan pengelasan yang tepat. Salah satu kelebihan utama Gradio ialah sokongannya terhadap prototaip pantas dan kebolehjelasan model. Menurut Abid et al. (2021), Gradio membolehkan pembangun "mencipta dan berkongsi antara muka untuk model pembelajaran mesin mereka dengan mudah, sekali gus meningkatkan ketelusan dan kebolehhasilan." Ini bermaksud model bukan sahaja boleh diuji dalam persekitaran pembangunan, malah boleh digunakan untuk demonstrasi, ujian pengguna, atau program pendidikan. Selain itu, satu lagi kelebihan Gradio ialah sokongannya terhadap maklum balas masa nyata, yang membantu dalam pengesahan model menggunakan input dunia sebenar. Ia membolehkan pengguna bukan teknikal memahami cara model bertindak balas terhadap pelbagai jenis imej, seterusnya memupuk kepercayaan dan kebolehfahaman dalam penyelesaian berasaskan kecerdasan buatan (AI) (Chien et al., 2022).

Secara keseluruhannya, integrasi Gradio meningkatkan kepraktisan model pengesanan buli siber ini dengan membolehkan pengujian interaktif, pengelasan imej secara langsung, dan mengurangkan halangan teknikal untuk penggunaan oleh pelbagai lapisan pengguna.

4.0 HASIL

4.1 Keputusan Model dicadangkan

Bahagian ini memberikan gambaran menyeluruh mengenai eksperimen yang melibatkan model asas dan model yang dicadangkan bagi mendapatkan pemahaman yang lebih jelas tentang kelebihan dan kekurangan algoritma yang digunakan. Matlamat eksperimen ini adalah untuk memperoleh output accuracy yang lebih tinggi daripada model yang dicadangkan berbanding model asas semasa. Seterusnya, hasil eksperimen akan dianalisis untuk mendapatkan kesimpulan. Eksperimen ini melibatkan beberapa fasa, di mana fasa pertama bermula dengan pengujian model-model terhadap set data yang berbeza. Model asas (EfficientNetB0-SVM) dan model yang dicadangkan (EfficientNetB3-SVM-XGBoost) diuji dari segi prestasi dan metrik klasifikasi menggunakan dua set data. Seterusnya, model yang dicadangkan dinilai melalui eksperimen ablation.

Jadual 1 Ringkasan set data

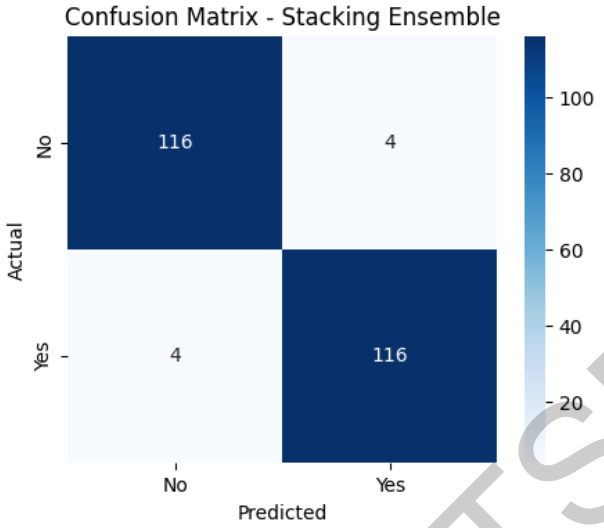
Label Imej	Dataset 1 (dataset asas)	Dataset 2 (dataset GitHub)
1 “Buli Siber”	600 imej	600 imej
0 “Bukan Buli Siber”	600 imej	600 imej
Jumlah	1200 imej	1200 imej

Model yang dicadangkan iaitu EfficientNetB3-SVM-XGBoost telah diuji menggunakan kedua-dua dataset untuk merekod prestasi dan accuracy bagi tujuan perbandingan model. Dataset tersebut diperbahagi terlebih dahulu kepada dua subset, iaitu set latihan (80%) dan set ujian (20%), sebagaimana prosedur yang konsisten dengan fasa eksperimen sebelum ini. Bagi Dataset 1, model yang dicadangkan mengklasifikasikan dengan betul 232 daripada 240 imej ujian, manakala 8 imej diklasifikasikan secara salah, seperti yang diperlihatkan dalam confusion matrix pada Rajah 12. Nilai purata accuracy model meningkat daripada 94.58% kepada 96.67%.

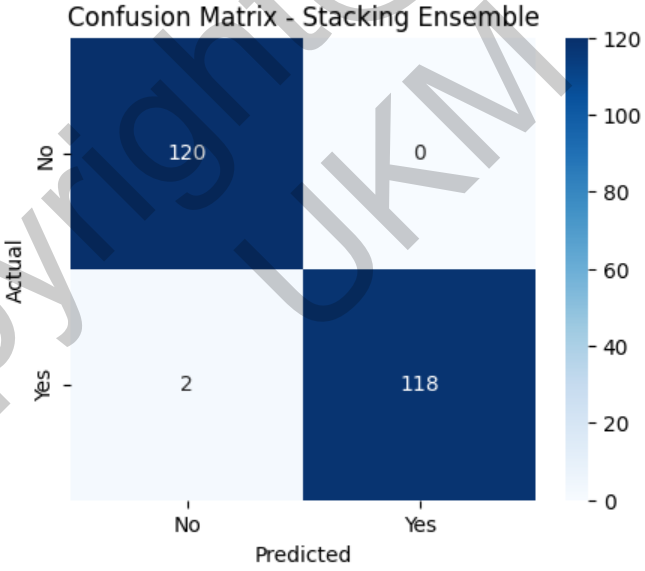
Model EfficientNetB3-SVM-XGBoost mencapai nilai yang konsisten bagi semua metrik seperti precision, recall dan F1-score, di mana semua metrik tersebut seimbang dan sama sebanyak 97% untuk kedua-dua kelas ("Ya" dan "Tidak"). Sebagai perbandingan, precision bagi kelas "Tidak" dalam model ini lebih tinggi berbanding model asas, menunjukkan bahawa model yang dipertingkatkan mempunyai keupayaan yang tinggi dalam menghasilkan keputusan positif yang tepat bagi kedua-dua kelas.Selain itu, nilai recall yang sama tinggi iaitu 97% untuk kedua-dua kelas "Ya" dan "Tidak" menggambarkan bahawa model ini lebih konsisten dalam mengenal pasti semua kejadian sebenar sama ada imej buli siber atau bukan buli siber berbanding model asas.

Secara sebaliknya, F1-score model yang dicadangkan juga lebih tinggi daripada model asas yang memberi gambaran tentang keseimbangan yang lebih baik antara precision dan recall. Purata macro F1-score dan weighted F1-score masing-masing mencatatkan nilai 97%, satu peningkatan sedikit berbanding model asas yang hanya mencapai 95%. Seterusnya untuk Dataset 2, model yang dicadangkan mengklasifikasikan dengan betul 238 daripada 240 imej ujian, dan hanya 2 imej diklasifikasikan secara salah, seperti yang ditunjukkan dalam confusion matrix pada Rajah 4-6. Purata accuracy model mencapai 99.17%, yang merupakan keputusan tertinggi setakat eksperimen dijalankan. Model ini juga mencapai precision sebanyak 100% bagi kelas "Ya" dan 98% bagi kelas "Tidak", menunjukkan bahawa tidak terdapat sebarang kes false positive dalam kelas "Ya". Nilai recall yang dicatatkan sebanyak 98% menunjukkan bahawa model hampir dapat mengenal pasti kes buli siber sebenar dengan sempurna. Bagi kelas "Tidak", recall mencapai 100%, menandakan bahawa semua imej bukan buli siber berjaya dikenal pasti dengan betul.Selain itu, F1-score untuk kedua-dua kelas "Ya" dan "Tidak" adalah sebanyak 99%, yang mengesahkan kestabilan dan keberkesanan model dalam menangani keseimbangan antara precision dan recall. Purata macro F1-score dan weighted F1-score masing-masing memperoleh skor yang memberangsangkan sebanyak 99%. Peningkatan daripada 97% kepada 99% menunjukkan bahawa model yang dipertingkatkan lebih cekap dalam mengendalikan imej buli siber dan bukan buli siber tanpa memihak kepada mana-mana kelas, sekali gus mengesahkan peningkatan dalam keadilan pengelasan (class-wise fairness) dan generalisasi model.

Eksperimen ablasi merupakan satu pendekatan sistematik yang digunakan untuk menganalisis sumbangan setiap komponen dalam sesuatu model, sistem atau reka bentuk eksperimen dengan cara menyingkirkan atau mengubah suai elemen-elemen tertentu. Teknik ini sering digunakan dalam bidang pembelajaran mesin, pembelajaran mendalam dan penyelidikan saintifik untuk menilai sejauh mana sesuatu ciri atau pengubahsuaian memberi kesan terhadap prestasi keseluruhan sistem. Dalam konteks model yang dicadangkan, beberapa komponen tambahan seperti MinMaxScaler dan Principal Component Analysis (PCA) telah digabungkan bagi meningkatkan keupayaan model asas, dengan matlamat untuk mencapai prestasi optimum, keteguhan (robustness) dan kebolehan pengamuman (generalization) model. Namun demikian, kajian terperinci adalah penting bagi menentukan sejauh mana komponen-komponen tersebut mempengaruhi prestasi model dalam mengklasifikasikan imej buli siber dengan tepat.Sehubungan itu, beberapa eksperimen ablasi telah dijalankan dengan tujuan untuk menganalisis secara sistematik kepentingan dan sumbangan setiap ciri dalam model EfficientNetB3-SVM-XGBoost.



Rajah 3 *confusion matrix* bagi dataset 1



Rajah 4 *confusion matrix* bagi dataset 2

Jadual 5 Ringkasan eksperimen ablasi

Experiment	Feature Involved
Ablasi Pengekstrakan Ciri	Penggantian EfficientNetB3 dengan EfficientNetB0
Ablasi Pengekstrakan Ciri	Penyingkiran EfficientNetB3 (penggunaan piksel mentah)
Ablasi Pemprosesan Ciri	Penyingkiran MinMaxScaler
Ablasi Pemprosesan Ciri	Penyingkiran CLAHE
Ablasi Klasifikasi-Meta	Penggunaan SVM sahaja
Ablasi Klasifikasi-Meta	Penggunaan XGBoost sahaja

Jadual 6 Laporan klasifikasi bagi model EfficinetNetB3-SVM-XGBoost sebelum ablasi

Dataset 1	**Accuracy**	**Precision**	**Recall**	**F1-Score**
	96.67%			
Kelas “Ya”		97%	97%	97%
Kelas “Tidak”		97%	97%	97%
Macro average				97%
Weighted average				97%

Jadual 7 Laporan klasifikasi bagi ablasi EfficientNetB3

	Accuracy	**Precision**	**Recall**	**F1-Score**
EfficientNetB3	96.67%			
EfficientNetB0	95.83%			
Kelas “Ya”		98%	93%	96%
Kelas “Tidak”		94%	98%	96%
Macro average				96%
Weighted average				96%

Jadual 8 **Laporan klasifikasi bagi ablasi pengekstrakan ciri**

	Accuracy	Precision	Recall	F1-Score
Pengekstrakan ciri (EfficientNetB3)	96.67%			
Piksel mentah	95.42%			
Kelas “Ya”		94%	97%	95%
Kelas “Tidak”		97%	94%	95%
Macro average				95%
Weighted average				95%

Jadual 9 **Laporan klasifikasi bagi ablasi MinMaxScaler**

	Accuracy	Precision	Recall	F1-Score
Dengan MinMaxScaler	96.67%			
Tanpa MinMaxScaler	92.50%			
Kelas “Ya”		93%	93%	93%
Kelas “Tidak”		93%	93%	93%
Macro average				93%
Weighted average				93%

Jadual 10 **Laporan klasifikasi bagi ablasi CLAHE**

	Accuracy	Precision	Recall	F1-Score
Dengan CLAHE	96.67%			
Tanpa CLAHE	93.33%			
Kelas “Ya”		95%	92%	93%
Kelas “Tidak”		92%	95%	93%
Macro average				93%
Weighted average				93%

Rajah 11 Laporan klasifikasi bagi ablasi klasifikasi meta (SVM)

	Accuracy	Precision	Recall	F1-Score
Ridge classifier	96.67%			
SVM only	95.83%			
Kelas “Ya”		94%	97%	96%
Kelas “Tidak”		97%	94%	96%
Macro average				96%
Weighted average				96%

Rajah 12 Laporan klasifikasi bagi ablasi klasifikasi meta (XGBoost)

	Accuracy	Precision	Recall	F1-Score
Ridge classifier	96.67%			
XGBoost only	92.50%			
Kelas “Ya”		93%	92%	92%
Kelas “Tidak”		92%	93%	93%
Macro average				92%
Weighted average				92%

5.0 KESIMPULAN

Secara keseluruhannya, kajian ini berjaya membangunkan satu model pengesanan buli siber berasaskan imej yang terenhans, yang mengatasi model asas dari segi ketepatan dan kekukuhan prestasi. Dengan menaik taraf pengekstrak ciri daripada EfficientNetB0 kepada EfficientNetB3, serta mengintegrasikan teknik pra-pemprosesan lanjutan seperti CLAHE, MinMaxScaler, dan PCA, disamping penggunaan pembelajaran ensembel jenis stacking yang menggabungkan SVM dan XGBoost dengan bimbingan meta-pengelasan Ridge, model ini telah mencapai peningkatan prestasi pengelasan yang ketara. Model terenhans ini menunjukkan ketepatan melebihi 99%, berbanding prestasi lebih rendah yang dicapai oleh model asas, sekali gus mengesahkan nilai tambah bagi setiap penambahbaikan seni bina yang diperkenalkan.Selain itu, eksperimen ablasi yang dijalankan turut memberikan bukti empirikal mengenai sumbangan setiap ciri terhadap kejayaan keseluruhan model.

Dari sudut pandang yang lebih luas, perbincangan kajian ini turut menekankan implikasi dunia sebenar dalam usaha memerangi buli siber di platform media sosial, serta menggariskan kepentingan pengesanan awal melalui penyelesaian berasaskan AI.Walaupun terdapat beberapa keterbatasan seperti saiz set data yang terhad dan isu salah klasifikasi isyarat (gesture), model yang dibangunkan ini menawarkan asas yang kukuh untuk pelaksanaan praktikal. Penyelidikan masa hadapan boleh menumpukan kepada penggunaan set data yang lebih besar, lebih pelbagai, serta pembangunan ke arah pelaksanaan masa nyata (real-time implementation).

Walaupun model ini menunjukkan ketepatan dan prestasi yang tinggi apabila diuji merentasi kedua-dua set data, masih terdapat ruang untuk penambahbaikan pada masa hadapan. Salah satu cadangan utama ialah menggabungkan pembelajaran multimodal (multi-modal learning), di mana data visual dan teks (seperti kapsyen imej atau komen) digunakan bersama-sama untuk memberikan konteks yang lebih baik dalam mengesan buli siber. Pendekatan ini mampu mengurangkan salah klasifikasi yang berpunca daripada kekaburan maklumat visual semata-mata. Walau bagaimanapun, pendekatan ini tidak dapat dilaksanakan dalam kajian ini kerana keperluan keupayaan perkakasan komputer yang tinggi.Selain itu, integrasi mekanisme perhatian (attention mechanisms) atau vision transformers (ViTs) boleh diterokai bagi membolehkan model memberi tumpuan yang lebih berkesan pada kawasan penting dalam imej, menghasilkan keputusan yang lebih kontekstual. Set data juga boleh diperluaskan dengan lebih banyak sampel dunia sebenar, termasuk variasi dalam pencahayaan, latar belakang, dan gerak isyarat yang lebih halus, untuk mengurangkan kecondongan model dan meningkatkan generalisasi. Tambahan lagi, teknik pengoptimuman model seperti quantization dan pruning boleh diterapkan untuk menghasilkan model yang lebih ringan dan sesuai untuk pelaksanaan masa nyata pada peranti berkuasa rendah. Pendekatan keterjelasan model (explainability techniques) seperti Grad-CAM juga disarankan untuk meningkatkan ketelusan dan kepercayaan pengguna terhadap keputusan yang dihasilkan. Model pengesanan buli siber yang dicadangkan ini mempunyai potensi kuat untuk digunakan secara nyata, khususnya pada platform media sosial seperti Instagram, Facebook, TikTok, dan Twitter yang berpusatkan kandungan visual. Integrasi model ini boleh membantu sistem moderasi automatik untuk menandakan imej berpotensi berbahaya secara masa nyata, sekaligus membantu moderator manusia menguruskan jumlah besar kandungan yang dihasilkan pengguna. Pengesanan awal buli siber adalah penting untuk mencegah kesan psikologi, pengasingan sosial, dan tekanan emosi jangka panjang, khususnya dalam kalangan remaja.

6.0 PENGHARGAAN

Pertama sekali, saya ingin mengucapkan kesyukuran saya kepada Tuhan kerana dengan izin-Nya, saya dapat menyiapkan laporan teknikal untuk memenuhi syarat Ijazah Sarjana Muda Sains Komputer dengan Kepujian dengan sempurna dalam tempoh masa yang ditetapkan. Selain itu, saya turut bersyukur kerana memberikan saya kesihatan fizikal dan mental yang baik serta memberikan kesabaran untuk menghadapi segala masalah dan cabaran sepanjang persiapan projek ini. Saya juga ingin merakamkan jutaan terima kasih kepada penyelia projek tahun akhir saya, Dr Wandeep kaur A/P Ratan Singh atas segala bimbingan, dorongan, nasihat dan kritikan yang amat berharga sepanjang perjalanan menyiapkan usulan ini. Saya juga mengucapkan terima kasih yang tidak terhingga kerana saya mungkin tidak dapat menyiapkan projek ini dengan jayanya tanpa bantuan beliau. Sekalung penghargaan dan terima kasih saya ucapkan kepada ibu saya, Lachumi A/P Manikam, serta ahli keluarga lain atas sokongan dan dorongan yang diberikan dalam membantu saya menyiapkan projek ini. Ucapan terima kasih yang tidak terhingga juga ditujukan kepada rakan-rakan seperjuangan atas tunjuk ajar dan bantuan yang telah diberikan sepanjang pelaksanaan projek ini. Saya juga ingin mengucapkan terima kasih kepada setiap individu yang telah membantu saya, sama ada secara sedar atau tidak, dalam menjayakan projek ini.

7.0 RUJUKAN

Avila-Garzon, C., Bacca-Acosta, J., Duarte, J., & Betancourt, J. (2021). Augmented Reality in Education: An Overview of Twenty-Five Years of Research. Contemporary Educational Technology, 13(3).

Buijtendijk, M. F., Barnett, P., & van den Hoff, M. J. (2020, March). Development of the human heart. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics (Vol. 184, No. 1, pp. 7-22)*. Hoboken, USA: John Wiley & Sons, Inc.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining" gamification". *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9-15).

Elmqaddem, N. (2019). Augmented reality and virtual reality in education. Myth or reality?. *International journal of emerging technologies in learning*, 14(3).

Guntur, M. I. S., Setyaningrum, W., Retnawati, H., & Marsigit, M. (2020, January). Assessing the potential of augmented reality in education. *Proceedings of the 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning* (pp. 93-97).

Gonzalez, A. A., Lizana, P. A., Pino, S., Miller, B. G., & Merino, C. (2020). Augmented reality-based learning for the comprehension of cardiac physiology in undergraduate biomedical students. Advances in Physiology Education, 44(3), 314-322.

Ahmad, J., & Meerah, S. (2002). Pemupukan budaya penyelidikan di kalangan guru di sekolah: Satu penilaian. Bangi: Universiti Kebangsaan Malaysia.

Johnson, L., Levine, A., Smith, R., & Stone, S. (2010). The 2010 Horizon Report. New media consortium. 6101 West Courtyard Drive Building One Suite 100, Austin, TX 78730. Horizon.

Li, W., Grossman, T., & Fitzmaurice, G. (2014, April). CADament: a gamified multiplayer software tutorial system. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3369-3378).

Lund, A. M. (2001). Measuring usability with the use questionnaire12. Usability interface, 8(2), 3-6.

Manzano-León, A., Camacho-Lazarraga, P., Guerrero, M. A., Guerrero-Puerta, L., Aguilar-Parra, J. M., Trigueros, R., & Alias, A. (2021). Between level up and game over: A systematic literature review of gamification in education. *Sustainability, 13*(4), 2247.

Nuanmeesri, S., Kadmateekarun, P., & Poomhiran, L. (2019). Augmented Reality to Teach Human Heart Anatomy and Blood Flow. *Turkish Online Journal of Educational Technology-TOJET*, *18*(1), 15-24.

O'Shea, P., & Scapin, T. (2020). A Review of Commercially Available Educational Augmented Reality Apps. *Innovate Learning Summit,* 251-261.

Osadchyi, V. V., Valko, N. V., & Kuzmich, L. V. (2021, March). Using augmented reality technologies for STEM education organization. *Journal of Physics: Conference Series* (Vol. 1840, No. 1, pp. 012027). IOP Publishing.

Pallant, J. (2007). SPSS Survival Manual: A Step-By-Step Guide To Data Analysis Using SPSS for Windows. *CrowsNest West: Allen & Unwin.*

Ramli, R. Z., Marobi, N. A., & Ashaari, N. S. (2021). Microorganisms: Integrating augmented reality and gamification in a learning tool. *International Journal of Advanced Computer Science and Applications, 12*(6).

Roopa, D., Prabha, R., & Senthil, G. A. (2021). Revolutionizing education system with interactive augmented reality for quality education. *Materials Today: Proceedings, 46*, 3860-3863.

Rozhenko, O. D., Darzhaniya, A. D., Bondar, V. V., & Mirzoian, M. V. (2021). Gamification of education as an addition to traditional educational technologies at the university. In CEUR Workshop Proceedings (Vol. 2914, pp. 457-464).

Saleem, A. N., Noori, N. M., & Ozdamli, F. (2021). Gamification applications in E-learning: a literature review. Technology, Knowledge and Learning, 1-21.

Vidal-Balea, A., Blanco-Novoa, Ó., Fraga-Lamas, P., & Fernández-Caramés, T. M. (2021). Developing the next generation of augmented reality games for pediatric healthcare: an open-source collaborative framework based on arcore for implementing teaching, training and monitoring applications. *Sensors, 21*(5), 1865.

Zhao, F. (2019). Using Quizizz to Integrate Fun Multiplayer Activity in the Accounting Classroom. *International Journal of Higher Education*, *8*(1), 37-43.

Kavin Arasan A/L
Mudiarasan (A195661) Dr.
Wandeep Kaur A/P Ratan
Singh
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia