



Theme
**Artificial Intelligence and Data Science:
Prospects and Future Applications**

2024 5th International Conference on Artificial Intelligence and Data Sciences (AiDAS)



CONFERENCE PROCEEDINGS

All rights reserved. Copyright ©2024 by IEEE.
IEEE Catalog Numbers: ISBN 979-8-3315-2855-3 Part Number CFP24AIG-ART

2024 5th International Conference on Artificial Intelligence and Data Sciences (AiDAS)

CONFERENCE PROCEEDINGS

All rights reserved.
Copyright ©2024 by IEEE.

IEEE Catalog Numbers:
ISBN 979-8-3315-2855-3
Part Number CFP24AIG-ART

2024 5TH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND DATA SCIENCES (AIDAS)

© 2024 by IEEE. All rights reserved.

Copyright and Reprint Permission:

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For reprint or republication permission, email to IEEE Copyrights Manager at pubs-permissions@ieee.org. All rights reserved. Copyright ©2024 by IEEE.

IEEE Catalog Numbers

ISBN 979-8-3315-2855-3

Part Number CFP24AIG-ART

Table of Contents

	<i>page</i>
Organising Committee	xii
Reviewers	xiii
Author Index	xvi
 Enhancing Threat Detection in Network Security Based on Augmented Feature Chai Guo Qiang, Aun Yichiet, Png Wen Hao, Goh Hock Guan	 1
 Exploring Random Forest Regression for Financial Distress Detection Guo Dong Hou, Dong Ling Tong, Soung Yue Liew, Peng Yin Choo	 7
 Interactive Parallel Coordinates Visualization Platform for Survey Data M. Bakri, Afrina N.A., M. Hamiz	 13
 Development of a Comprehensive Dashboard for Monitoring Student Performances Chow Wai Kean, Fatimah Audah Md. Zaki, Kok-Chin Khor	 18
 Optimizing Heart Disease Prediction: Integrating Relief and Logistic Regression for Feature Selection Tan Jia Qi, Fatimah Audah Md. Zaki	 24
 Lecturers' and Students' Perspectives on Using Chat-GPT in Academics for Creative Problem Solving: A Dilemma or an Opportunity for Improvement? Samirah Razali	 30
 Consistent Fuzzy Preference Relation Method To Evaluate The Weightage For Malay Translated Hadith Expert Judgment Shaiful Bakhtiar bin Rodzman, Ku Muhammad Naim Ku Khalif	 35
 Identifying Ripeness in Chokanan Mango Fruit Using K-Nearest Neighbor Nur Athirah Syafiqah Noramli, Raihana Md Saidi, Hajar Izzati Mohd Ghazalli	 42
 Text Normalization of Penang Hokkien Dialect leveraging Adapted Soundex Algorithm Yu Liang Lai, Yen Min Jasmina Khaw, Seng Poh Lim, Tien Ping Tan	 48
 Ambidexterity and High-Performance Work Systems: The Roles of Knowledge Sharing, Leadership Style, and Social Media Saide Saide, Nurfitria Ningsi, Didi Muwardi, Dian Jelita, Nita Antasari	 55
 Exploring the Relationship between eWOM and Purchase Intention through Machine Learning Algorithms Bhawana Kothari, Ambica Prakash Mani, V. M. Tripathi	 60

Table of Contents (continued)

Modelling and Forecasting Mortality Rates Using a Neural Network-Integrated Lee-Carter Model	65
Nurul Aityqah Yaacob, Ibrahim Mohamed, Siti Noor Dina Ahmad	
A Review : DDoS Attack Detection using Clustering Algorithms	70
Amit Kumar Mishra, Siddhant Thapliyal, Junedh Siddiqui, Rohit Bhatt, Keshav Naithani, Ankit Joshi	
Adaptive Reconstruction Methods Selection Framework for Compressive Sampling	75
Ke Xin Chew, Boon Yaik Ooi, Xin Yi Kh'ng, Woan Lin Beh	
Analysis of the influence of Green Artificial Intelligence on Agricultural and biological fields using bibliometrics.	81
Jonitha Anand, Nooraini Yusof, Hadhrami Ab Ghani, Kiran Kumar Thoti	
Reduced Noise SMOTE in Machine Learning Model: Application in Water Quality Classification with Imbalanced Datasets	87
Norashikin Nasaruddin, Nurulkamal Masseran, Wan Mohd Razi Idris, Ahmad Zia Ul-Saufie	
A Review on Deep Learning and Hybrid Model for Forecasting Residential and Commercial Buildings Energy Consumption	93
Siti Sarah Mat Isa, Ahmad Asri Abd Samat, Nur Hazahsha Shamsudin, Mohd Najib Mohd Hussain, Siti Salwa Mat Isa, Abdul Malek Saidina Omar	
Oral Ulcer Classification System Using CNN	99
Ahmad Ikmal Naim Mod Azam, Azliza Mohd Ali, Wannoraini Abdul Latif, Wan NurHazirah Wan Ahmad Kamil, Mukarramah Zainal, Nurzeatul Hamimah Abdul Hamid	
A Real-time System for Monitoring Student Drowsiness in the Classroom using the Deep Learning Model YOLOv8	105
Nur Hannani Hanis Mohd Hamidi, Nor Afirdaus Zainal Abidin, Raihah Aminuddin, Chew Chiou Sheng, Khyrina Airin Fariza Abu Samah, Siti Diana Nabilah Mohd Nasir	
MyDAS Corpus: Malay Social Media Texts for Detecting Depression, Anxiety, and Stress on Facebook	111
Zaaba Ahmad, Azlinah Mohamed, Mike Conway, Rozanizam Zakaria, Noor Hasimah Ibrahim Teo, Ruhaila Maskat	
Enhancing Mental Health Detection in Malay Social Media: A Comparative Study of RNN Architectures and Attention Mechanisms	121
Zaaba Ahmad, Azlinah Mohamed, Mike Conway, Ruhaila Maskat	
Combination of Boruta's algorithm and lasso logistic regression in establishing blood-based biomarkers for Alzheimer's disease	132
Mohammad Nasir Abdullah, Yap Bee Wah	
Exploring Feature Selection Methods in Predicting College Students' Churn in Telecom Services	138
Jiejun Li, Norriza Hussin, Yaozu Luo, Yanhua Zhong	
Utilizing the Mol2Vec algorithm to extract features of HIV drug interactions.	144
Felix Indra Kurniadi, Risma Yulistiani	

Table of Contents (continued)

Handling Sparsity Problem using Removal and Imputation in the HUPA-UCM Diabetes Type 1 Dataset	149
Felix Indra Kurniadi, Risma Yulistiani	
Posture image recognition using a skeleton-based CNN model with late channel attention	154
Haochen Han, Nufei Cheng, Jing Tian	
Vision-based posture change detection using a self-attention guided CNN model	158
Nufei Cheng, Haochen Han, Jing Tian	
Sports Video Classification Using Convolutional Neural Network (CNN) with Normalization Flow	162
Nurul A. Emran, Nurul Izrin Md Saleh, Muhamad Zaidi Mohd Ali	
Crop Pest and Diseases Classification using ResNet and Inception Network	168
Tan Yanli, Azliza Mohd Ali, Sharifalillah Nordin, Wang Jin, Li Guoqin	
Performance Evaluation of Shariah-compliant Portfolios using Grey Relational Clustering and Markowitz Model	175
Farah Azaliney Mohd Amin, Md Nizam Udin, Huda Nabilah Binti Zulkefli, Izzah Nursyafiqah Binti Che Halim, Siti Norsyahilda Binti Ahmad Shafie, Siti Nazifah Zainol Abidin	
Measuring the Performance of Rice Production in Kuala Nerang region using Grey Relational Analysis	180
Farah Azaliney Mohd Amin, Nur Fathiah Fatin Mohamad Fauzi, Zahari Md. Rodzi	
A Study of Sentiment Analysis in Customer Food Preferences Through Machine Learning	185
Noryanti Nasir, Siti Meriam Zahari, Siti Sarah Januri, Nur Iffa Batrisyia Mohd 'Aidil, Nur Izzatun Nawa Mohd Pauzi, Anis Zahirah Zubir	
Stoppage to the Implementation of Internet of Things (IoT) in Construction Projects: from Lens of Civil Engineers in the United Kingdom	191
Nadzirah Zainordin, Salmaliza Salleh, Nafiya Farsana, Sui Lai Khoo, Ahmad Faris Omar, Putu Doddy Heka Ardana	
Performance Analysis of Support Vector Machine (SVM) for Diabetes Disease Detection	203
Ricky Krisdianto, Ivana Apriani, Hossey Masada, Hidayaturrahman	
From Clicks to Class Participation: Demystifying Student Engagement Factors with Attribute Ranking and Predictive Analytics	208
Rozita Hanapi, Mohd Norhisham Razali, Syazleena Yahya, Sharidatul Akma Abu Seman, Suhaida Halamy, Emelia Abdul Rahim	
Improved Layered Architecture: Integration of IoT with Blockchain for Increased Security and Scalability	214
Muhammad Jahangir, Ming Jie Lee, Ban-Hoe Kwan, Tariq Kamal, Tong-Yuen Chai, Wing Son Loh	

Table of Contents (continued)

Pneumonia Detection System Using Convolutional Neural Network	221
Muhammad Fazrul Izzat, Siti 'Aisyah Sa'dan, Ummu Fatimah Mohd Bahrin, Siti Nurul Hatie Ishak, Sharifah Nurulhikmah Syed Yasin, Normalina Ibrahim @ Mat Nor	
The Development of Instruments to Measure Students Behavioural Intention towards Adopting Artificial Intelligence (AI) Technologies in Educational Settings	227
Roszi Naszariah Nasni Naseri, Jhanghiz Syahrivar, Izni Syamsina Saari, Wan Kalthom Yahya, Geetha Muthusamy	
A Novel Method For Studying Mosquito Oviposition Behaviour Using Computer Vision And Deep Learning Algorithm	233
Song-Quan Ong, Pradeep Isawasan, Gomesh Nair, Khairulliza Ahmad Salleh	
Detection of Viral Pneumonia in Radiographic Images Using Deep Learning	239
Wasakorn Laesanklang, Phonlapee Lohajareekul	
An integrated model for predicting student achievement efficiency using Data Envelopment Analysis and Genetic Programming approach	245
Nor Faezah Mohamad Razi, Suraya Masrom, Norhayati Baharun, Anis Zafirah Azmi	
Investigation Of Data Leakage In Stroke Lesion Detection Through Diffusion Weighted Imaging Using Data Splitting Approaches	256
Jasmine Wang Thye Wei, Hang See Pheng, Ong Kok Haur	
Tinira ni Benny: A Recipe Recommender System to Minimize Food Waste	262
Avian Lance Basco, Priscilla Mariah Licup, Darren Andrei Longno, Michelle Andrea Martinez, Angel Lou Yabut, Norshuhani Zamin	
AI in Action: Unveiling Factors Influencing AI and Human Collaboration in Malaysian Academic Libraries	268
Fadhilah Aman, Norizan Zakaria	
Code Mixing Malay Online Text Sentiment Analysis: Proposed Feature Extraction and Feature Selection Scheme	274
Surendran Selvaraju, Nilam Nur Amir Sjarif, Mohd Syahid Mohd Anuar	
Sentiment Analysis in Code Mixing Malay Text: A Review	280
Surendran Selvaraju, Nilam Nur Amir Sjarif, Mohd Syahid Mohd Anuar	
Assessing the Impact of Artificial Intelligence Adoption among Academicians in Higher Education Institutions in Malaysia: A Theoretical Model	286
Nur Farhanis Ahmad Anuar, Nur Balqishanis Zainal Abidin, Sarah A'fifah Abdullah Sani, Siti Nurlaili Karim, Nur Amalina Mat Jan	
Real-Time Personal Protective Equipment Compliance Detection using You Only Look Once	292
Nurhanisah Aminuddin, Nor Azuana Ramli, Agus Pratondo	
Probability Distribution Model for Flood Prediction in the Trolak River, Perak	298
Tiong Sze Yen, Nur Amalina Mat Jan, Nur Balqishanis Zainal Abidin, Loshini Thiruchelvam, Noraini Ibrahim, Zanariah Zainudin, Muhammad Fadhil Marsani	

Table of Contents (continued)

GPT-4 Powered Virtual Analyst for Fundamental Stock Investment by Leveraging Qualitative Data	304
Teck Wan Teo, Soung-Yue Liew, Chee Henn Chng, Jing Ying Ng, Zong Zheng Khoo	
Data Quality Frameworks: A Systematic Review	310
Ismail El Gayar, Hesham Hassan, Khaled T. Wassif	
Machine Learning Model for Predicting Community Integrity towards Solid Waste Management and Environmental Care	316
Faridah Zulkipli, Nurul Husna Jamian, Suraya Masrom, Khairul Irwan Khalid, Zulkifli Mohd Nopiah, Nittaya Kerdprasop	
Evaluating Back Translation and Misspelling Correction Utilization on Indonesian AES	322
Elvina Amadea Tanaka, Steven Christian, Anderies, Andry Chowanda	
Temporal Shifts in Customer Sentiment: Analysing Airasia Pre- And Post-Pandemic Reviews	327
Sheema Liza Idris, Masurah Mohamad	
Interval-Valued Fuzzy Bézier Surface Approximation	332
Nor Ashikin Sahrom, Mohammad Izat Emir Zulkifly, Siti Nur Idara Rosli	
An Unsupervised Machine Learning Approach for Heart Disease Prediction	337
Wing Son Loh, Yu Jiun Lim, Lloyd Ling, Ren Jie Chin, Chen Kang Lee, Ming Jie Lee	
Intrusion Detection System (IDS) Classifications using Hyperparameter Tuning for Machine Learning and Deep Learning	344
Tan May May, Zanariah Zainudin, Norliana Muslim, Nurul Syafidah Jamil, Nur Amalina Mat Ja, Noraini Ibrahim, Nor 'Afifah Binti Sabri	
Token-Based Path Allocation for High Throughput and Scalability in Indoor Autonomous Mobile Robot Systems	350
Zong Zheng Khoo, Chee-Siang Wong, Soung-Yue Liew, Teck Wan Teo, Jing Ying Ng, Chee Henn Chng	
A Hybrid ResNet - MobileNet Deep Learning Model for Smart Bin Waste Classification	356
Nur Zulaikha Binti Azhar, Noor Hasimah Ibrahim Teo, Raseeda binti Hamzah, Rosniza binti Roslan, Ruhaila Maskat	
Comparative Analysis of ARIMA and LSTM Approaches for Monthly River Flow Forecasting in Terengganu	363
Noraini Ibrahim, Norhaiza Ahmad, Nur Amalina Mat Jan, Zanariah Zainudin, Nurul Syafidah Jamil, Ammar Azlan	
Finite Order Universal Portfolio With Moment-Generating Function And Machine Learning Model	369
Sheong Wei Phoon, You Chuen Chong, Wing Son Loh, Kee Seng Kuang, Gillian Yi Han Woo	
The Discovery and Solution of the Pseudo-Reasoning Issue for Constructing Cost-Effective Multi-Agent Frameworks in Large Language Models	376
Jing Ying Ng, Soung-Yue Liew, Chee Henn Chng, Teck Wan Teo, Zong Zheng Khoo	

Table of Contents (continued)

Random Forest-Based Approach To Android Malware Detection In Machine Learning	381
Muhammad Hafiz Mohd Saufan, Samsiah Ahmad, Zalikha Zulkifli, Nur Hasni Nasarudin, Mohamed ImranMohamed Ariff, Nor Aslily Sarkam	
Benign and Malignant Detection and Classification for Small Size Image of Breast Tumor Recognition System using U-Net Model	386
Suryanti Awang, Saumya Kumar, Nur Syafiqah Mohd Nafis, Raihanah Haroon	
Sentiment Analysis on YouTube Comment Data for the Candidate Debate in the 2024 Presidential Election of the Republic of Indonesia	392
Kartika Purwandari, Meganingrum Arista Jiwanggi, Evi Yulianti	
Application Integration Framework for Large Language Models	398
Joe Ee HO, Boon Yaik OOI, Markus WESTNER	
Malaysian Residential Property: A Forecasting Model	404
Nur Atiqah Khalizan, Nur Alia Farhana Muhammad Harith Fadzillah, Norani Amit	
Application of Artificial Neural Network for Accurate Prediction of Photocatalytic Dye Degradation	410
Ameer Shuhaib Azmil, Mohd Fadhil Majnis, Mohd Azam Mohd Adnan, Syahriza Ismail	
AI-Driven Animation Creation: Application Exploration and Potential Risks	416
Li Yixuan, Azahar Harun,Rao Dongyu	
Optimizing Personalized Recommendation in E-services Platforms Using AI	422
Kai Yi Lim, Ana Nabilah Sa'uadi, Nur Athirah Nabila Mohd Idros, Nurul Syafidah Jamil	
An Analysis on Cost and Income Prediction System using Multiple Linear Regression and Business Intelligence for Entrepreneur in Fruit Crop	430
Mohd Taufik Mishan, Aimi Liyana Amir, Nurhashikin Mohd Salleh	
OYEN: A User-Centric LLM-Based Bilingual Healthcare Chatbot	435
Owen Chin, Nurul Syafidah Jamil, Zanariah Zainudin, Nor Azizah Hitam, Noraini Ibrahim, Ahmad Hakimi Ahmad Sa'ahiry	
Patient Attitude on the Application of Artificial Intelligence in Diabetes Care	441
Maslin Masrom, Logeswary Krisnan, Yazriwati Yahya, Manimaran Krishnan Kaundan	
Redefining Activation Dynamics: An Empirical Comparison of Oscillating, Mixture, and Adaptive Activation Functions in Deep Reinforcement Learning	446
Liong Gele, Tomas Maul	
PetVision: Improved MobileViTv3-based Image Classification for Cat Breed Identification	452
Jackson Ang, Shannie Tannaris, Andry Chowanda, Anderies Anderies	
Determining The Readiness Factors To Adapt Blockchain Technology In The Palm Oil Industry To Facilitate Circular Economy Transition	457
Zurida Ishak, Ahmad Shazwan Ahmad Amran, Mohammad Dalvi Esfahani, Ramesh Kumar Ayyasamy, Meilina Ong Bt. Abdullah @ Ong Li Mei, Yeoh Jo Ann	

Table of Contents (continued)

Myocardial Ischemia Classification Through Ecg Using Deep Learning Hazimah Ahmad Shafawi, Sharifalillah Nordin	463
Knowledge Visualization of Student Performance Based on Attendance by using Prescriptive Analytic Shakiroh Khamis, Mazida Ahmad, Ahmad Subhi Bin Zolkafly, Soong Hoong Cheng, Hapini Awang, Nur Suhaili Mansor	467
Artificial Neural Network as the Classifier of the Muscle Contraction Pattern during Stepping Activity Aiman Syabil Mazlan, Sofea Nur Batrisyia Suhaizan, Zariq Hishamudin, Nor Aziyatul Izni, Siti Nor Aini Mohd Aslam, Sarah 'Atifah Saruchi	474
GLCM-Based Feature Extraction and CNN Approach for Fruit Freshness Detection Mas Azza Abu Bakar, Zalikha Zulkifli, Masurah Mohamad, Samsiah Ahmad	480
Preliminary Study on the Accessibility and Learning Experience of Low-Code Development Platforms Jit Theam Lim, Chen Kang Lee, Ean Heng Lim, Pei Voon Wong, King Tak Yew, Chek Yee Ooi	486
Risk of Cardiovascular Heart Disease Using Data Mining Approach Nur Alya Ahmad Shafri, Nur Afiqah Aliya Mohd Nizam, Siti Aida Sheikh Hussin, Zalina Zahid	492
Fuzzification and Defuzzification of Intuitionistic Fuzzy B-Spline Curve Approximation Model for Visualization of Complex Uncertainty Data Siti Nur Idara Rosli, Mohammad Izat Emir Zulkifly	498
Exploring Academicians Behavioral Intention Towards Blended Learning System: ANN Approach Nor Aslily Sarkam, Nor Hazlina Mohammad, Nur Izzah Jamil, Nor Faezah Mohamad Razi, Samsiah Ahmad, Ahmad Nur Azam Ahmad Ridzuan	504
Augmented Reality (AR) for Food Safety: Implementing an AR Inspection App with a Machine Learning Model Nicholas Ngiam Li Xuan, Saw Seow Hoon, Saw Seow Hui	510
The Factors Associated Inadequacy Retirement Fund Among Employee Provident Fund (EPF) Members Rabiatul Adawiyah Zainudin, Nur Aina Izureen Mohd Amir Izwan, Muhammad Hilmi Samian, Anas Fathul Ariffin, Siti Nurasyikin Shamsuddin, Diana Juniza Juanis	516
Bio-Inspired Algorithm for Text Classification: Comparison between Ant Colony Algorithm and Olex-GA Ahmad Nazmi Fadzal, Nazuha Fadzal, Ahmad Nazri Fadzal, Norlina Mohd Sabri, Mazidah binti Puteh	522
Fuzzy Topsis with Ratings based on Sub-Criteria for Selection of Supplier Samsiah Abdul Razak, Nazirah Ramli, Muhammad Ikmal Hakim Nor Azmi, Muhammad Ikmal Mohd Nor, Hadizul Alif Hakimi Noreddie	528
An application of machine learning on social media for solid waste management Siti Sarah Januri, Nadira Afiqa binti Mohd Rozi, Noryanti Nasir, Siti Natasha Mustaf-Jab, Nurul Natasha Sahar	533

Table of Contents (continued)

Deep Learning-Based Car Plate Number Recognition (CPR) in Videos Stream	538
Siti Nurezzaty Rafek, Siti Nur Kamaliah Kamarudin, Yuzi Mahmud	
Reimagine Sentiments in Nature Language through Crystalize-Water Visualization	544
Wen-Hao Png, Fang-Yu Chang, Jie-Jin Lo, Su-Chu Hsu	
Explainable Convolutional Neural Network for Corn Disease Classification	549
Mark Cholo Rabadon, Nicole Cathlene Astrologo, Mary Joy Lapitan, Marita Pinili, Ranzivelle Marianne Roxas-Villanueva	
Sentiment Analysis on Food Delivery Services in Malaysia	555
Aisya Azhar, Masurah Mohamad, Zalikha Zulkifli, Nurulhuda Zainuddin, Itaza Afiani Mohtar	
Implementation of Light Gradient Boosting Machine (LGBM) for Customer Services Feedback Classification	561
Maria Susan Anggreainy	
Product Label Clustering Based on Search Keywords for Product Name Recommendations in Ecommerce	566
Bryan, Daniel Alexander, Maria Susan Anggreainy, Devi Fitriana, Ajeng Wulandari	
Chinese Monolingual Sentiment Classifier for Social Media Data using Corpus-based Approach	570
Nur Atiqah Sia Abdullah, Shanell Crisenn Low Cheng Cheng, Marshima Mohd Rosli	
Defending Against Digital Threats: Machine Learning Techniques for Cyber Persecution Detection	575
P. Chinnasamy, Ramesh Kumar Ayyasamy, Kesavan Krishnan, Vijay Kumar, Norazira Binti A Jalil, Ajmeera Kiran	

Author Index

A

Abdul Malek Saidina Omar 93
 Afrina N.A. 13
 Agus Pratondo 292
 Ahmad Asri Abd Samat 93
 Ahmad Faris Omar 191
 Ahmad Hakimi Ahmad Sa'ahiry 435
 Ahmad Ikmal Naim Mod Azam 99
 Ahmad Nazmi Fadzal 522
 Ahmad Nazri Fadzal 522
 Ahmad Nur Azam Ahmad Ridzuan 504
 Ahmad Shazwan Ahmad Amran 457
 Ahmad Subhi Bin Zolkafly 467
 Ahmad Zia Ul-Saufie 87
 Aiman Syabil Mazlan 474
 Aimi Liyana Amir 430
 Aisya Azhar 555
 Ajeng Wulandari 566
 Ajmeera Kiran 575
 Ambica Prakash Mani 60
 Ameer Shuhaib Azmil 410
 Amit Kumar Mishra 70, 75
 Ana Nabilah Sa'uadi 422
 Anas Fathul Ariffin 516
 Anderies 322
 Anderies Anderies 452
 Andry Chowanda 322, 452
 Angel Lou Yabut 262
 Anis Zafirah Azmi 245
 Anis Zahirah Zubir 185
 Ankit Joshi 70
 Aun Yichiet 1
 Avian Lance Basco 262
 Azahar Harun 416
 Azlinah Mohamed 111, 121
 Azliza Mohd Ali 99, 168

B

Ban-Hoe Kwan 214
 Bhawana Kothari 60
 Boon Yaik OOI 398
 Bryan 566

C

Chai Guo Qiang 1
 Chee Henn Chng 304, 350, 376
 Chee-Siang Wong 350
 Chek Yee Ooi 486
 Chen Kang Lee 337, 486
 Chew Chiou Sheng 105
 Chow Wai Kean 18

D

Daniel Alexander 566
 Darren Andrei Longno 262
 Devi Fitriana 566
 Diana Juniza Juanis 516
 Dian Jelita 55
 Didi Muwardi 55
 Dong Ling Tong 7

E

Ean Heng Lim 486
 Elvina Amadea Tanaka 322
 Emelia Abdul Rahim 208
 Evi Yulianti 392

F

Fadhilah Aman 268
 Fang-Yu Chang 544
 Farah Azaliney Mohd Amin 175, 180
 Faridah Zulkipli 316
 Fatimah Audah Md. Zaki 18, 24
 Felix Indra Kurniadi 144, 149

Organizer



Co-Organizer



Technical Co-Sponsor



Author Index (continued)**G**

Geetha Muthusamy 227
 Gillian Yi Han Woo 369
 Goh Hock Guan 1
 Gomesh Nair 233
 Guo Dong Hou 7

H

Hadhrami Ab Ghani 81
 Hadizul Alif Hakimi Noreddie 528
 Hajar Izzati Mohd Ghazalli 42
 Hang See Pheng 256
 Haochen Han 154, 158
 Hapini Awang 467
 Hazimah Ahmad Shafawi 463
 Hesham Hassan 310
 Hidayaturrahman 203
 Hossey Masada 203
 Huda Nabilah Binti Zulkefli 175

I

Ibrahim Mohamed 65
 Ismail El Gayar 310
 Itaza Afiani Mohtar 555
 Ivana Apriani 203
 Izni Syamsina Saari 227
 Izzah Nursyafiqah Binti Che Halim 175

J

Jackson Ang 452
 Jasmine Wang Thye Wei 256
 Jhanghiz Syahrivar 227
 Jie-Jin Lo 544
 Jiejun Li 138
 Jing Tian 154, 158
 Jing Ying Ng 304, 350, 376
 Jit Theam Lim 486
 Joe Ee HO 398
 Jonitha Anand 81
 Junedh Siddiqui 70

K

Kai Yi Lim 422
 Kartika Purwandari 392
 Kee Seng Kuang 369
 Kesavan Krishnan 575
 Keshav Naithani 70
 Khairul Irwan Khalid 316
 Khairulliza Ahmad Salleh 233
 Khaled T. Wassif 310
 Khyrina Airin Fariza Abu Samah 105
 King Tak Yew 486
 Kiran Kumar Thoti 81
 Kok-Chin Khor 18
 Ku Muhammad Naim Ku Khalif 35

L

Li Guoqin 168
 Liong Gele 446
 Li Yixuan 416
 Lloyd Ling 337
 Logeswary Krisnan 441
 Loshini Thiruchelvam 298

M

Manimaran Krishnan Kaundan 441
 Maria Susan Anggreainy 561, 566
 Marita Pinili 549
 Mark Cholo Rabadon 549
 Markus WESTNER 398
 Marshima Mohd Rosli 570
 Mary Joy Lapitan 549
 Mas Azza Abu Bakar 480
 Maslin Masrom 441
 Masurah Mohamad 327, 480, 555
 Mazida Ahmad 467
 Mazidah binti Puteh 522
 M. Bakri 13
 Md Nizam Udin 175
 Meganingrum Arista Jiwanggi 392
 Meilina Ong Bt. Abdullah @ Ong Li Mei 457
 M. Hamiz 13



Author Index (continued)

- Michelle Andrea Martinez 262
 Mike Conway 111, 121
 Ming Jie Lee 214, 337
 Mohamed Imran Mohamed Ariff 381
 Mohammad Dalvi Esfahani 457
 Mohammad Izat Emir Zulkifly 332, 498
 Mohammad Nasir Abdullah 132
 Mohd Azam Mohd Adnan 410
 Mohd Fadhil Majnis 410
 Mohd Najib Mohd Hussain 93
 Mohd Norhisham Razali 208
 Mohd Syahid Mohd Anuar 274, 280
 Mohd Taufik Mishan 430
 Muhamad Zaidi Mohd Ali 162
 Muhammad Fadhil Marsani 298
 Muhammad Fazrul Izzat 221
 Muhammad Hafiz Mohd Saufan 381
 Muhammad Hilmi Samian 516
 Muhammad Ikmal Hakim Nor Azmi 528
 Muhammad Ikmal Mohd Nor 528
 Muhammad Jahangir 214
 Mukarramah Zainal 99
- N**
- Nadira Afiqa binti Mohd Rozi 533
 Nadzirah Zainordin 191
 Nafiya Farsana 191
 Nazirah Ramli 528
 Nazuha Fadzal 522
 Nicholas Ngiam Li Xuan 510
 Nicole Cathlene Astrologo 549
 Nilam Nur Amir Sjarif 274, 280
 Nita Antasari 55
 Nittaya Kerdprasop 316
 Nooraini Yusof 81
 Noor Hasimah Ibrahim Teo 111, 356
 Nor 'Afifah Binti Sabri 344
 Nor Afirdaus Zainal Abidin 105
 Noraini Ibrahim 298, 344, 435
 Norani Amit 404
 Norashikin Nasaruddin 87
 Nor Ashikin Sahrom 332
 Nor Aslily Sarkam 381, 504
 Norazira Binti A Jalil 575
 Nor Aziyatul Izni 474
 Nor Azizah Hitam 435
 Nor Azuana Ramli 292
 Nor Faezah Mohamad Razi 245, 504
 Norhayati Baharun 245
 Nor Hazlina Mohammad 504
 Norizan Zakaria 268
 Norliana Muslim 344
 Norlina Mohd Sabri 522
 Normalina Ibrahim @ Mat Nor 221
 Norriza Hussin 138
 Norshuhani Zamin 262
 Noryanti Nasir 185, 533
 Nufei Cheng 154, 158
 Nur Afiqah Aliya Mohd Nizam 492
 Nur Aina Izureen Mohd Amir Izwan 516
 Nur Alia Farhana Muhammad Harith Fadzillah 404
 Nur Alya Ahmad Shafri 492
 Nur Amalina Mat Ja 344
 Nur Amalina Mat Jan 286, 298
 Nur Athirah Nabila Mohd Idros 422
 Nur Athirah Syafiqah Noramli 42
 Nur Atiqah Khalizan 404
 Nur Atiqah Sia Abdullah 570
 Nur Balqishanis Zainal Abidin 286, 298
 Nur Farhanis Ahmad Anuar 286
 Nur Fathiah Fatin Mohamad Fauzi 180
 Nurfitria Ningsi 55
 Nurhanisah Aminuddin 292
 Nur Hannani Hanis Mohd Hamidi 105
 Nurhashikin Mohd Salleh 430
 Nur Hasni Nasarudin 381
 Nur Hazahsha Shamsudin 93
 Nur Iffa Batrisyia Mohd 'Aidil 185
 Nur Izzah Jamil 504
 Nur Izzatun Nawa Mohd Pauzi 185
 Nur Suhaili Mansor 467
 Nur Syafiqah Mohd Nafis 386
 Nurul A. Emran 162
 Nurul Aityqah Yaacob 65
 Nurulhuda Zainuddin 555
 Nurul Husna Jamian 316
 Nurul Izrin Md Saleh 162



Author Index (continued)

Nurulkamal Masseran 87
 Nurul Natasha Sahar 533
 Nurul Syafidah Jamil 344, 422, 435
 Nurzeatul Hamimah Abdul Hamid 99
 Nur Zulaikha Binti Azhar 356

O

Ong Kok Haur 256
 Owen Chin 435

P

P. Chinnasamy 575
 Pei Voon Wong 486
 Peng Yin Choo 7
 Phonlapee Lohajareekul 239
 Png Wen Hao 1
 Pradeep Isawasan 233
 Priscilla Mariah Licup 262
 Putu Doddy Heka Ardana 191

R

Rabiatul Adawiyah Zainudin 516
 Raihah Aminuddin 105
 Raihanah Haroon 386
 Raihana Md Saidi 42
 Ramesh Kumar Ayyasamy 457, 575
 Ranzivelle Marianne Roxas-Villanueva 549
 Rao Dongyu 416
 Raseeda binti Hamzah 356
 Ren Jie Chin 337
 Ricky Krisdianto 203
 Risma Yulistiani 144, 149
 Rohit Bhatt 70
 Rosniza binti Roslan 356
 Roszi Naszariah Nasni Naseri 227
 Rozanizam Zakaria 111
 Rozita Hanapi 208
 Ruhaila Maskat 111, 121, 356

S

Saide Saide 55
 Salmaliza Salleh 191
 Samirah Razali 30
 Samsiah Abdul Razak 528
 Samsiah Ahmad 381, 480, 504
 Sarah A'fifah Abdullah Sani 286
 Sarah 'Atifah Saruchi 474
 Saumya Kumar 386
 Saw Seow Hoon 510
 Saw Seow Hui 510
 Seng Poh Lim 48
 Shaiful Bakhtiar bin Rodzman 35
 Shakiroh Khamis 467
 Shanell Crisenn Low Cheng Cheng 570
 Shannie Tannaris 452
 Sharidatul Akma Abu Seman 208
 Sharifah Nurulhikmah Syed Yasin 221
 Sharifalillah Nordin 168, 463
 Sheema Liza Idris 327
 Sheong Wei Phoon 369
 Siddhant Thapliyal 70
 Siti Aida Sheikh Hussin 492
 Siti 'Aisyah Sa'dan 221
 Siti Diana Nabilah Mohd Nasir 105
 Siti Meriam Zahari 185
 Siti Natasha Mustaf-Jab 533
 Siti Nazifah Zainol Abidin 175
 Siti Noor Dina Ahmad 65
 Siti Nor Aini Mohd Aslam 474
 Siti Norsyahilda Binti Ahmad Shafie 175
 Siti Nurasyikin Shamsuddin 516
 Siti Nurezzaty Rafek 538
 Siti Nur Idara Rosli 332, 498
 Siti Nur Kamaliah Kamarudin 538
 Siti Nurlaili Karim 286
 Siti Nurul Hatie Ishak 221
 Siti Salwa Mat Isa 93
 Siti Sarah Januri 185, 533
 Siti Sarah Mat Isa 93
 Sofea Nur Batrisyia Suhaizan 474
 Song-Quan Ong 233



Author Index (continued)

Soong Hoong Cheng 467
 Soung Yue Liew 7
 Soung-Yue Liew 304, 350, 376
 Steven Christian 322
 Su-Chu Hsu 544
 Suhaida Halamy 208
 Sui Lai Khoo 191
 Suraya Masrom 245, 316
 Surendran Selvaraju 274, 280
 Suryanti Awang 386
 Syahriza Ismail 410
 Syazleena Yahya 208

T

Tan Jia Qi 24
 Tan May May 344
 Tan Yanli 168
 Tariq Kamal 214
 Teck Wan Teo 304, 350, 376
 Tien Ping Tan 48
 Tiong Sze Yen 298
 Tomas Maul 446
 Tong-Yuen Chai 214

U

Ummu Fatimah Mohd Bahrin 221

V

Vijay Kumar 575
 V. M. Tripathi 60

W

Wang Jin 168
 Wan Kalthom Yahya 227
 Wan Mohd Razi Idris 87
 Wannoraini Abdul Latif 99
 Wan NurHazirah Wan Ahmad Kamil 99
 Wasakorn Laesanklang 239
 Wen-Hao Png 544
 Wing Son Loh 214, 337, 369

Y

Yanhua Zhong 138
 Yaozu Luo 138
 Yap Bee Wah 132
 Yazriwati Yahya 441
 Yen Min Jasmina Khaw 48
 Yeoh Jo Ann 457
 You Chuen Chong 369
 Yu Jiun Lim 337
 Yu Liang Lai 48
 Yuzi Mahmud 538

Z

Zaaba Ahmad 111, 121
 Zahari Md. Rodzi 180
 Zalikha Zulkifli 381, 480, 555
 Zalina Zahid 492
 Zanariah Zainudin 298, 344, 435
 Zariq Hishamudin 474
 Zong Zheng Khoo 304, 350, 376
 Zulkifli Mohd Nopiah 316
 Zurida Ishak 457



MyDAS Corpus: Malay Social Media Texts for Detecting Depression, Anxiety, and Stress on Facebook

1st Zaaba Ahmad

*College of Computing,
Informatics & Mathematics
Universiti Teknologi MARA
Cawangan Perak Kampus Tapah
Tapah, Perak, Malaysia
zaaba.ahmad@uitm.edu.my*

2nd Azlinah Mohamed

*YSL Strategic Alliance
Shah Alam, Selangor, Malaysia
azlinahmohamed@gmail.com*

3rd Mike Conway

*Centre for Digital Transformation of Health
The University of Melbourne
Melbourne, Victoria, Australia
mike.conway@unimelb.edu.au*

4th Rozanizam Zakaria

*Department of Psychiatry
Kulliyah of Medicine
International Islamic University Malaysia
Kuantan, Pahang, Malaysia
rozanizam@iiu.edu.my*

5th Noor Hasimah Ibrahim Teo

*College of Computing,
Informatics & Mathematics
Universiti Teknologi MARA,
Cawangan Melaka Kampus Jasin
Merlimau, Melaka, Malaysia
hasimah.iteo@uitm.edu.my*

6th Ruhaila Maskat*

*College of Computing,
Informatics & Mathematics
Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
ruhaila256@uitm.edu.my*

Abstract—The application of Natural Language Processing (NLP) in mental health monitoring has significantly expanded; however, the specific challenges of interpreting Depression, Anxiety, and Stress (DAS) in Malay language social media texts have not been adequately addressed. This gap underscores the need for NLP solutions that are sensitised to the linguistic and cultural specificities of Malay-speaking populations. This study develops and validates a specialised Malay language corpus from social media content, targeting DAS. Utilising a hybrid ground truth strategy that integrates self-reports with expert assessments, the research offers methodological refinements in the analysis of Malay linguistic patterns and the deployment of machine learning classifiers to efficiently identify mental health indicators. The paper reviews existing methodologies, outlines a novel corpus development strategy, and discusses classifier performance. The Decision Tree classifier achieved the highest F1 score of 0.75, followed by the Support Vector Machine (SVM) with an F1 score of 0.73, and Random Forest with 0.70. Multinomial Naive Bayes (MNB) and K-Nearest Neighbors (KNN) demonstrated lower performances with F1 scores of 0.55 and 0.52 respectively. Comprehensive analyses using bi-gram networks and t-SNE visualisations explore the nuanced linguistic indicators of mental health states, culminating in a discussion of the implications for future NLP applications in mental health monitoring.

Index Terms—Mental Health Detection, Natural Language Processing, Malay Language, Corpus Development, Social Media, Depression, Anxiety, Stress, Decision Tree, Support Vector Machine

I. INTRODUCTION

Mental health issues have garnered increasing global attention as a critical aspect of the Sustainable Development Goals (SDGs) [1], emphasising the importance of health and

well-being for all. These international objectives underscore the urgent need for effective detection, early intervention, and continuous support in mental health. The research focuses on creating and utilising a specialised Malay language corpus to apply NLP techniques for identifying mental health indicators, addressing significant gaps in the broader context of mental health research.

Social media platforms, where individuals frequently express thoughts and feelings, provide a rich data source that can be analysed to detect early signs of mental health conditions. This capability is crucial for enabling healthcare professionals to offer timely assistance to those in need. However, the application of NLP for mental health detection in the Malay language is still in its nascent stages [2]. The unique linguistic and cultural nuances of the Malay language can significantly influence the expression and interpretation of mental health conditions, necessitating tailored NLP solutions.

The study aims to bridge this gap by developing a specialised Malay language corpus from social media, focusing on Depression, Anxiety, and Stress (DAS). This corpus uses a hybrid ground truth strategy that combines self-reports with expert assessments to enhance data reliability and validity. Challenges such as dataset imbalance, symptom overlap across different conditions, and the intricacies of multiclass classification are addressed to refine the detection process. This research aligns with the SDGs by enabling more culturally and linguistically appropriate interventions, thus improving global mental health resources. The study's practical implications could lead to more effective and timely interventions for



Fig. 1: Generic methodology for corpus development in predicting DAS

mental health conditions in the Malay-speaking population.

The paper is structured as follows: section 2 reviews related works, issues, and challenges while section 3 describes the methodology, detailing the data collection and labelling processes. Section 4 explores the reliability and validity of the corpus and section 5 presents a detailed corpus examination. Section 6 evaluates the effectiveness of traditional machine learning classifiers while section 7 reflects on the implications of the findings. Finally, Section 8 concludes the paper with a summary of the contributions to the field and outlines suggestions for future research directions.

II. RELATED WORKS, ISSUES, AND CHALLENGES

Research into applying NLP for mental health detection in the Malay language has been limited, despite extensive investigations in other domains such as sentiment analysis, information retrieval, sarcasm, and emotion detection [2]. This study addresses this gap by analysing social media textual data to uncover patterns indicative of mental health conditions, aiming for early detection and intervention.

Various studies have leveraged alternative signals to discern positive and negative DAS indicators in mental health corpus development. Positive indicators might include self-reports of conditions, medication use, or suicide attempts, while negative indicators could be the absence of such reports or expressions of well-being. This study lays the foundation by establishing a robust ground truth strategy before data collection, aligning with studies validating DAS's presence through inter-rater reliability tests, such as Cohen's kappa analysis after corpus development. Figure 1 presents a generic approach to corpus development for predicting DAS, inspired by [5].

This framework guides various strategies for effectively identifying these indicators within social media content. Five basic strategies for predicting DAS in social media scenarios are considered to determine the ground truth. The most common strategy is self-disclosure, where individuals assert that they experience a particular condition or display DAS-consistent behaviours through phrases like 'I was diagnosed with...'. Positive annotations include self-reports of conditions, medication use, suicide attempts, and self-stated stress.

Another strategy involves using DAS screening tools in corpus creation. Many studies utilise screening tools and questionnaires, such as the Center for Epidemiologic Studies Depression Scale (CES-D), the Depression Anxiety Stress Scale (DASS-21), the Beck Depression Inventory (BDI) and the Patient Health Questionnaire (PHQ). The third strategy involves creating dictionaries or lexicons from corpora through keyword search. Most studies search for symptom words or disorder names in social media posts or profiles, viewing community involvement as a DAS indicator. This examination includes formal community participation on SMPs like Weibo [16], [17], Reddit [18], [19], Facebook [20], [21], and LiveJournal [22].

Additionally, many studies rely on self-annotations to determine mental health status on social media, often involving domain specialists annotating content using diagnostic criteria or through clinical interviews. Crowdsourcing services are also frequently used for data collection. The collaboration between computer scientists and domain experts is particularly significant in ensuring data accuracy. Some researchers adopt mixed methodologies, combining different strategies to establish the ground truth of their corpora [5]. This comprehensive approach often includes collecting self-reported data, verified with DAS levels using various screening tools and questionnaires.

A significant challenge in implementing NLP for mental health detection in Malay text is the scarcity of comprehensive, high-quality datasets tailored for this purpose. Given the sensitive nature of the data, creating such datasets must navigate ethical considerations and respect for privacy [3]. Additionally, obtaining accurately labelled data representing DAS conditions prevalent among the Malay-speaking population proves to be a significant hurdle.

Bias in NLP models for mental health detection is another critical issue. These biases may arise from imbalanced datasets, algorithmic preferences, or pre-existing societal biases embedded within the data [4]. NLP models for mental health detection must also carefully consider the linguistic and cultural nuances of the Malay language [2]. Accurately interpreting and detecting mental health issues requires incorporating colloquialisms, dialects, and cultural references often unique to Malay.

The challenges are further compounded by the distinct linguistic characteristics of the Malay language, such as specific expressions, idiomatic phrases, and cultural references that generic NLP models, typically designed for English or other widely studied languages, need to capture. Without accommodating these language-specific nuances, there is a

significant risk of misinterpretations and inaccurate results. Furthermore, the limited availability of labelled data for mental health-related tasks in Malay restricts the development of robust and accurate detection tools.

Cultural context is incredibly significant in how mental health concerns are expressed and interpreted within Malay language data. Ensuring that NLP models are sensitive to these cultural dimensions is essential to prevent misinterpretations and biased outcomes. While much of the existing research has focused on binary classification tasks [5], the imperative to delve into multi-class classification for conditions such as DAS is growing. Developing models capable of identifying varying severities of these conditions is both a challenge and an opportunity. Moreover, ensuring that these models generalise effectively across different social media platforms and user demographics remains a persistent concern [6].

Developing automated or semi-automated tools to support practical mental health assessments necessitates collaboration between NLP experts, psychologists, and healthcare professionals [7]. Bridging the technical and clinical knowledge gaps is crucial for these models' accuracy and relevance. Ethical considerations, particularly in managing privacy, consent, and biases when analysing mental health data from social media, are paramount, as emphasised in various studies [8]–[12]. This paper addresses these multifaceted issues, paving the way for more accurate and culturally sensitive interventions for the Malay-speaking community. Overcoming these challenges underscores the necessity of a balanced approach to leveraging big data while ensuring ethical integrity in mental health computing.

III. METHODOLOGY

The methodology begins by selecting a suitable Social Media Platform (SMP). Despite the popularity of Twitter, Reddit, and Instagram, Facebook has been chosen for this research. Facebook's global reach and unique features make it a powerful platform for data mining and text classification in the mental health field [13], [14]. This decision is based on three main reasons: Facebook's high population-to-user ratio, the trust placed in it over other SMPs in adopting health-related information [15], and empirical observations that Malaysians frequently turn to Facebook's support groups for mental health assistance. Figure 2 illustrates the methodology for corpus development, outlining the stages involved in the process.

The high proportions of Malaysian Facebook users have made a large data pool invaluable for health researchers and data scientists. This extensive and continuously growing dataset of user experiences and emotional states allows for precise and thorough text classification and data mining, providing a comprehensive view of mental health patterns within the Malaysian demographic.

The work revealed that Facebook support groups play an integral role in the field of mental health, particularly in Malaysia. These online spaces are increasingly becoming a refuge for individuals seeking assistance, offering them a platform to share their mental health experiences and emotions

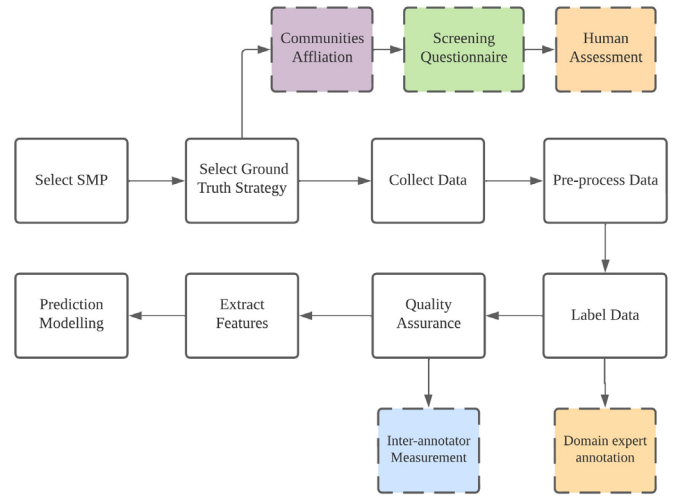


Fig. 2: Proposed hybrid methodology for corpus development in predicting DAS in the Malay language

openly. Such interactions within these groups contribute significantly to a rich reservoir of first-hand data, enabling text classification techniques to discern mental health trends and indicators by analysing the language used in these discussions. This makes Facebook an exceptional resource for the non-invasive detection of early mental health issues.

In the context of Facebook, trust among Malaysians is more prevalent within private groups. Their willingness to share personal information in these secure, enclosed spaces provides valuable data for effective data mining and text classification. It is suggested that Facebook is perceived to be more respectful of user privacy than other social networks, particularly in these private groups. This perceived respect for data privacy is crucial for enabling meaningful research while maintaining user privacy.

A. Ground Truth Approach

The research employs a hybrid approach to ground truth strategies, as illustrated in Figure 2. Initially, the focus was on a Malaysian Mental Health support group on Facebook, utilising a community affiliation strategy. Due to technical challenges in obtaining consent from the extensive user base, a new support group was created. Members from the main group were invited to join this new group, with assistance from psychologists involved in the original group. Participants were required to complete a screening questionnaire using the Malay version of the DASS-21 [23] and provide demographic information.

Explicit consent was obtained from each member to use their data in the research. This consent process was conducted meticulously to ensure participants knew their responses and demographic information would be collected and analysed. Of nearly 30,000 Facebook users in the original support group, 200 agreed to participate in the survey. This included the DASS-21 questionnaire, providing demographic details, and

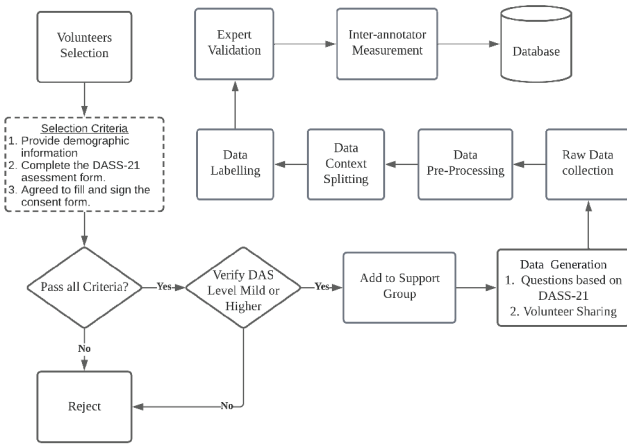


Fig. 3: Data collection flow diagram

granting consent. Of these 200 participants, only 160 met the eligibility criteria, showing at least a mild level of DAS. Those excluded either did not provide necessary information, had normal DAS levels, or were under the age requirement of 18 years.

B. Data Collection and Labelling

A support group comprising 160 individuals experiencing mental health issues was established, remaining active for three months. This group was facilitated by an in-house psychologist who led the discussions and provided professional advice to the participants. The initial phase of the group involved ice-breaking questions such as ‘How are you today?’, ‘Does anyone want to share anything?’ and ‘How did it all start?’. These questions helped ease participants into more open and meaningful conversations. The development and flow of these interactions, along with the study’s overall data collection process, are detailed in Figure 3, which presents a clear visual representation of the stages involved.

As the group progressed, the interactions became more organic, with members actively sharing their struggles and seeking advice on managing their conditions. To further guide these discussions, curated questions based on the DASS-21 questionnaire were introduced, specifically designed to elicit responses indicative of DAS signals. For example, the first question in the original DASS-21 questionnaire probes for the depression signal through expressions of sadness; thus, the curated questions had the same context or topic as the original question in the DASS-21 questionnaire. An in-house psychiatrist meticulously validated each curated question. This validation process was crucial to ensure that each response accurately reflected the specific DAS signal in question, thereby safeguarding the authenticity of the findings and minimising potential harm to the participants.

Approximately 100 active participants from the support group generated over 500 posts and comments. Active participants were defined based on their consistent engagement

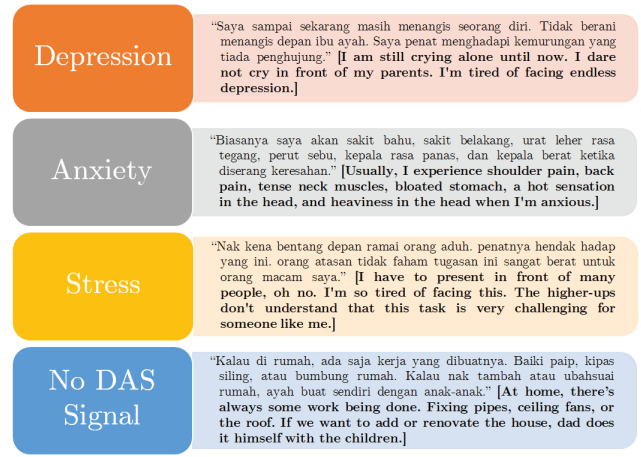


Fig. 4: Comparison of a simulated Malay DAS sample data and their English translations.

in the support group, evidenced by regular interactions with moderators and fellow members almost daily. The first step in processing this data involved extensive cleaning and pre-processing. As the analysis relied on embeddings, special characters were removed, a critical step given that most embeddings do not account for vector values of punctuation and other special characters. Furthermore, posts containing numbers, URLs, names, or identifiable information about the data owners were excluded, and all text was converted to lowercase. Another significant aspect of the text was the occasional use of mixed language or code-switching between Malay and English. To ensure consistency, all English terms were translated into their equivalent in the Malay language. This translation process also included the correction of typos and the expansion of abbreviations. For example, ‘utk’ was changed to ‘untuk’, ‘mrk’ to ‘mereka’, and ‘sgt’ to ‘sangat’.

To clarify the nature of the data samples and their translations, Figure 4 below compares the simulated Malay DAS sample data and their English translations. It is important to note that, following research guidelines for presenting social media data in publication [9] and the institutionally approved ethics protocol of the work, sensitive information is not revealed in this work. However, the simulated data closely resembles the original, providing an adequate representation for analysis and understanding.

A hybrid ground truth strategy for data labelling aligned with the methodology was employed. Due to the substantial data collected from the curated DAS questions, it was divided into two sets for manageability. Two expert psychiatrists labelled these sets, following the criteria in the original DASS21 questionnaire. Each psychiatrist reviewed one data set, manually labelling volunteers’ responses, especially those sharing experiences or seeking help. It was observed that individuals experiencing DAS often produced verbose posts. To address this, a method was developed where posts were contextually segmented based on DAS-specific signals or keywords identified by the experts. This approach enabled the

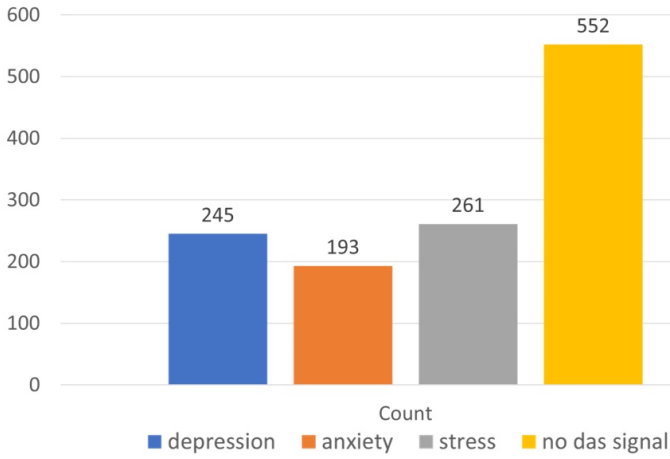


Fig. 5: The Malay DAS dataset

creation of a dataset more accurately reflecting the intended labels. Sentences without DAS indicators or irrelevant to DAS were labelled as ‘No DAS Signals,’ serving as negative data samples. Figure 5 showcases the final data tally, signifying its preparedness for subsequent processing.

IV. DATA VALIDATION

To ensure the precision and reliability of the data labelling framework, a rigorous inter-rater reliability analysis was conducted using Cohen’s Kappa statistic, as described by [24]. Cohen’s Kappa provides a measure of agreement between raters, considering the probability that agreement may occur by chance. The thresholds for interpreting Cohen’s Kappa generally range from values less than 0 indicating no agreement, to 0.01–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. The evaluation included two sets of data: Set 1 was labelled by Psychologist A and validated by Counsellors 1 and 2, and Set 2 was labelled by Psychologist B and validated by the same counsellors. This approach was designed to mitigate potential rater bias and confirm the consistency of expert evaluations.

The credibility of the counsellors in this analysis is notable. Both Counsellor 1 and Counsellor 2 possess extensive experience, each having served as a counsellor at a local university for approximately ten years. Additionally, both counsellors are certified clinical mental health counselling practitioners, enhancing their expertise and reliability in contributing to the inter-rater reliability analysis. This context strengthens the credibility of the data labelling process and highlights the depth of knowledge involved in ensuring the framework’s accuracy.

The Cohen’s Kappa evaluation yielded scores for Set 1 of 0.85 (Psychologist A vs Counsellor 1) and 0.86 (Psychologist A vs Counsellor 2), demonstrating strong agreement among raters. These scores indicate alignment in ratings and a harmonious interpretation of diagnostic criteria. Set 2, evaluated by Psychologist B, resulted in Cohen’s Kappa scores of 0.89

TABLE I: Cohen’s Kappa Scores for Different Rater Comparisons

Set	Psychologist	Counsellor	Cohen’s Kappa Score
Set 1	A	1	0.85
Set 1	A	2	0.86
Set 2	B	1	0.89
Set 2	B	2	0.90

(Psychologist B vs Counsellor 1) and 0.90 (Psychologist B vs Counsellor 2), continuing to show high-level agreement. This robust consistency across both sets underscores the impartiality of the assessment criteria and the precision of the raters’ evaluations.

The consistency in Cohen’s Kappa scores across both sets, detailed in Table I, reinforces the validity of the data labelling process. These scores, reflecting significant agreement amongst domain experts, solidify the foundation for subsequent analyses. The uniformity in the scores corroborates the diagnostic categories’ reliability, enhancing the conclusions’ integrity and robustness. Table I exemplifies the expertise and consensus of the raters, Psychologist A, Psychologist B, and the counsellors, in interpreting complex mental health conditions, highlighting their ability to maintain a high level of consistency in their evaluations.

V. CORPUS CHARACTERISTICS AND ANALYSIS

The corpus characteristics, detailed in Table II, provide a comprehensive overview of the dataset curated for analysing DAS indicators through social media interactions within a Malaysian context. The corpus comprises 510 posts, which include 270 responses to the DASS-21 questionnaire, 119 comments, and 121 instances of personal sharing. These elements collectively offer a rich dataset reflecting various facets of mental health discourse.

The dataset spans from 5 March 2022 to 4 June 2022, offering a snapshot of mental health discussions during this period. The average number of posts per group member is eight, indicating a moderately high level of engagement, which is crucial for deriving meaningful insights from social media data.

The length of the posts varies substantially, ranging from as few as three words to as many as 5091 words, with an average word count of 120. This post-length variance reflects how individuals express their mental states and interact within

TABLE II: Corpus Characteristics

Characteristic	Description
Total Number of Posts	510 Posts
Total Number of DASS-21 Response	270 Posts
Total Number of Comments	119 Posts
Total Number of Personal Sharing	121 Posts
Average Posts per Group Members	8 posts
Shortest Length of Posts	3 words
Longest Length of Posts	5091 words
Average Length of Posts	120 words
Data Collection Period	5 March 2022 - 4 June 2022

social media environments. Shorter posts often encapsulate concise, urgent expressions of mental states, while longer posts are likely to delve into detailed personal narratives or complex discussions about mental health challenges.

This diversity enhances the depth of the analysis. It poses unique challenges in natural language processing and text analysis, necessitating sophisticated techniques to interpret and classify such varied forms of expression accurately.

Compiling and meticulously categorising these data points, the corpus is a foundational element for deploying machine learning models to identify mental health indicators. It enriches the understanding of how mental health is articulated in a digital and multicultural landscape, enabling the exploration of nuanced linguistic and semantic patterns indicative of mental health states. This structured data collection and classification approach supports the development of targeted and effective mental health interventions based on social media interactions.

A. Corpus Analysis

The corpus analysis incorporates rigorous statistical and visual examinations to elucidate the language employed within the meticulously developed social media corpus. This analysis centres on bigram networks and t-SNE visualisations to identify and interpret the relationships and structures within the data, which are essential for refining machine learning models.

Bigram network visualisation illustrates the co-occurrence of terms within the corpus, underscoring the connections between frequently used words. This analysis is pivotal for identifying dominant themes and contextual relationships contributing to understanding the sentiment and topics discussed.

Furthermore, t-Distributed Stochastic Neighbor Embedding (t-SNE) transforms high-dimensional text data into a two-dimensional space. This method effectively visualises the clustering of posts based on linguistic features and semantic similarities, providing clear visual segmentation of data related to different mental health conditions.

These visualisations, generated through a Python script incorporating data cleaning, text vectorisation using scikit-learn's CountVectorizer, and dimensionality reduction with t-SNE, furnish qualitative and quantitative insights into the corpus characteristics. They elucidate the complex and diverse ways mental health concerns are articulated on social media platforms.

The insights garnered from these visual outputs are integral to the analysis. They enhance the understanding of the corpus and support the development of targeted mental health interventions based on social media interactions.

B. Bi-gram Network Analysis

The Bigram network analyses within the DAS corpus uncover distinct yet interconnected themes characterising each condition. Despite each condition's unique features, notable overlaps suggest common underlying mechanisms and shared influences.

The analysis across all conditions indicates significant discussions about deep emotional states. Terms such as 'rasa

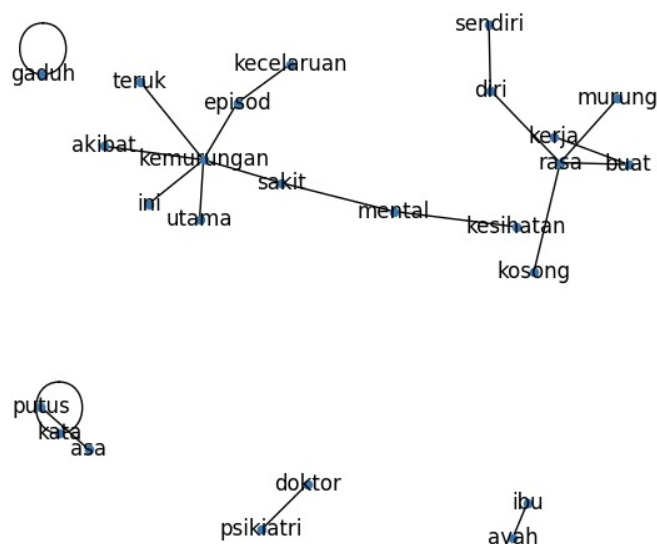


Fig. 6: Bigram network for depression highlighting key terms and their connections.

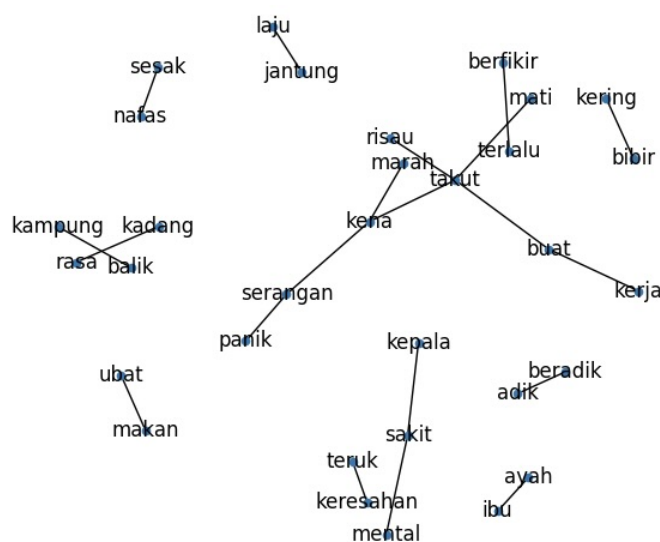


Fig. 7: Bigram network for anxiety illustrates the interplay of physiological and psychological terms.

'murung' (feel depressed), 'resah teruk' (severe anxiety), and 'susah hati' (heavy heart) are prevalent across the networks, highlighting the pervasive emotional distress associated with these conditions (refer to Figures 6, 7, and 8). The frequent mention of physical symptoms, such as 'sakit kepala' (headache) in stress discussions and 'jantung berdebar' (palpitating heart) in anxiety, underscores a psychosomatic overlap where emotional turmoil manifests as physical symptoms.

Discussions involving family members, such as 'ayah' (father) and 'ibu' (mother), are common across the conditions, indicating the significant role of familial relationships in influencing or exacerbating mental health conditions. Anxiety and stress mainly highlight the impact of social expectations and

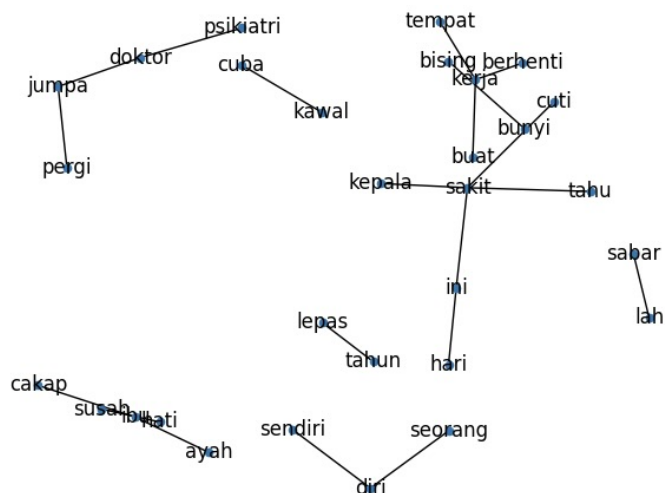


Fig. 8: Bigram network for stress showing environmental and immediate stressors.

obligations, such as ‘balik kampung’ (return to hometown) and workplace stress, showing how external pressures contribute to mental health issues. Coping strategies emerge as a recurring theme in discussions across all conditions, with advice such as ‘sabar kawal tahun lepas cuba’ (be patient, control, last year, try) reflecting common coping mechanisms.

Depression is uniquely characterised by its robust clinical framing within discussions, with terms like ‘kemurungan’ (depression) linked to ‘episod’ (episodes) and ‘teruk’ (severe), emphasising its episodic nature and clinical severity. Anxiety distinguishes itself with a focus on acute physiological reactions such as ‘serangan panik’ (panic attack), highlighting the immediacy of its symptoms and distinguishing it from the more pervasive and enduring symptoms discussed in depression and stress. Stress discussions often involve environmental factors and immediate stressors, such as ‘bunyi bising’ (noisy sound), which are less emphasised in depression and anxiety discussions.

This comprehensive analysis not only deepens the understanding of the specific characteristics of depression, anxiety, and stress but also highlights the shared themes and common coping strategies across these conditions. Recognising these overlaps is crucial for developing a holistic and accurate classification model that addresses the multifaceted nature of mental health disorders.

C. t-SNE Visualisation Analysis

The t-SNE visualisation is a tool for analysing the distribution and clustering of text data pertaining to DAS and texts lacking specific DAS signals. This section elaborates on the findings from the t-SNE analysis and correlates these insights with the preceding bi-gram network analysis.

The t-SNE plot in Figure 9 displays the general clustering of data points associated with anxiety (purple), depression (green), stress (blue), and no DAS signal (red). These distinct clusters illustrate how text data linked to each condition oc-

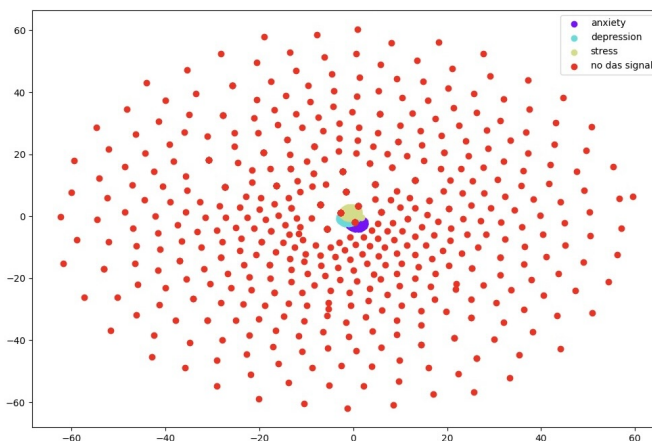


Fig. 9: t-SNE of text data by condition

cupies specific areas in the multidimensional space, revealing varying degrees of overlap among these areas.

The magnified view in Figure 10 uncovers intricate overlaps among the clusters. Notably, the anxiety and depression clusters show some intermingling, suggesting a linguistic or thematic commonality in how these conditions are articulated. This overlap might reflect shared symptoms or similar experiences in the textual data. Conversely, the stress cluster, while slightly overlapping with anxiety and depression, maintains a more distinct separation, indicating unique aspects of how stress is discussed.

Outliers, particularly those representing no DAS signal, are interspersed throughout the plot, highlighting the diversity in how non-specific mental health-related text data is portrayed. These outliers may contain unique linguistic features or less common symptom descriptions, providing valuable insights into atypical expressions of mental health conditions.

The t-SNE visualisation was created using a pre-processed high-dimensional dataset derived from the corpus to focus the analysis on DAS conditions. The pre-processing involved

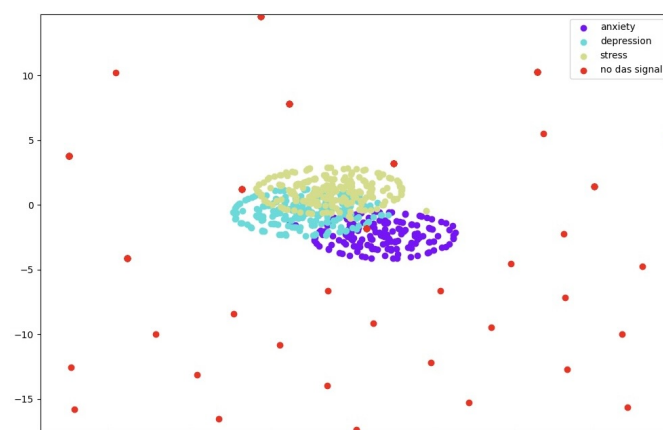


Fig. 10: Magnified t-SNE showing overlaps between DAS conditions

removing stopwords to minimise noise and employing bi-grams to capture more contextual information than single words alone. This approach offers a refined understanding of the text structure, emphasising the frequent connections between terms that may indicate mental health conditions.

The text data for each condition was vectorised using a CountVectorizer configured for bi-grams, and the high-dimensional vectors were then reduced to two dimensions using the t-SNE algorithm. This allowed for a visual depiction of the clustering tendencies of different mental health conditions. The t-SNE model was initialised with a random state to ensure reproducibility and applied to each condition's vectorised text data. The results were then visualised, using distinct colours for each condition to illustrate the clustering and overlaps clearly.

Integrating findings from the t-SNE analysis with the earlier bi-gram network analysis enhances the understanding of the linguistic landscape surrounding mental health discussions. While the bi-gram analysis directly linked key terms and their contextual associations, the t-SNE analysis spatially contextualises these connections within a multidimensional landscape. The consistency observed between the structural mappings in the bi-gram networks and the clustering in the t-SNE plots validates the robustness of the analytical methods.

This synergy between the two analyses facilitates a comprehensive exploration of semantic relationships and thematic groupings within the mental health discourse. It aids in the more informed development of detection and classification models for mental health conditions based on textual data. This integrated approach deepens theoretical understanding and significantly enhances practical applications in diagnosing and managing mental health issues.

VI. MODELLING WITH SYNTACTIC FEATURES AND COMMON ML CLASSIFIER

In this stage of the research, traditional machine learning classifiers are employed, including Support Vector Machine (SVM), Decision Trees, Multinomial Naive Bayes, and others, using syntactic features derived from TF-IDF. This foundational technique transforms text into a numerical representation, capturing the relative importance of words within and across documents. By quantifying word frequency while adjusting for word commonality, TF-IDF provides a normalised form of text data, which is essential for effective comparison and initial predictions. This baseline model serves as a benchmark, facilitating the evaluation of the effectiveness of more advanced neural network architectures and semantic features in subsequent research phases.

A. Common Classifier with MyDAS Data

To establish a baseline, multiple classifiers are employed to process data transformed through TF-IDF vectorisation, demonstrating the efficacy of traditional machine learning algorithms in classifying Malay DAS social media content. The following table presents the overall results of employing traditional ML classifiers with syntactic features, where each

TABLE III: Traditional ML Classifier with TF-IDF Performance Comparison

Classifier	F1 Score
Multinomial Naive Bayes (MNB)	0.55
K-Nearest Neighbors (KNN)	0.52
Support Vector Machine (SVM)	0.73
Random Forest	0.70
Decision Tree	0.75

classifier's performance is evaluated based on its F1 score. These scores provide insights into each classifier's effectiveness in handling multiclass classification challenges associated with DAS indications.

The results indicate that the Decision Tree model achieves the highest F1 score of 0.75, suggesting its strong suitability for the specific characteristics of the Malay DAS data due to its straightforward, hierarchical decision-making process. The Support Vector Machine (SVM), with an F1 score of 0.73, shows robust performance in high-dimensional spaces, such as those created by TF-IDF vectors, effectively capturing distinctions between different DAS states. Random Forest demonstrates good results with an F1 score of 0.70, benefiting from its ensemble approach that effectively mitigates the risk of overfitting and handles both bias and variance. Conversely, Multinomial Naive Bayes (MNB) and K-Nearest Neighbors (KNN) exhibit lower performance with F1 scores of 0.55 and 0.52, respectively, which may reflect limitations in handling the complex and nuanced features of the dataset.

B. ROC Curve Analysis of Classifiers in DAS Classification

The Receiver Operating Characteristic (ROC) curves offer a comprehensive view of each classifier's performance in distinguishing between DAS and No DAS conditions. This analysis is instrumental as it evaluates the trade-offs between true positive rates and false positive rates across various thresholds, providing insights into each classifier's robustness and reliability under different operational conditions.

Figure 11 illustrates the varied effectiveness of the classifiers, with the area under the curve (AUC) acting as a quantitative measure of each classifier's ability to effectively discriminate between classes. The Support Vector Machine (SVM) classifier exhibits the highest AUC of 0.94 ± 0.01 , signifying exceptional performance in differentiating between the classes. This performance is closely matched by Logistic Regression and Random Forest, with AUCs of 0.92 ± 0.01 and 0.93 ± 0.01 respectively, demonstrating significant discriminative power.

Although the Decision Tree classifier achieves the highest F1 score in the multiclass classification scenario, it exhibits a lower AUC of 0.87 ± 0.02 in the ROC analysis. This discrepancy may indicate a higher sensitivity to class imbalance or specific threshold settings in real-world applications. The K-Nearest Neighbors (KNN) classifier records the lowest AUC of 0.75 ± 0.01 , reflecting its challenges in handling the high-dimensional space of TF-IDF vectors and the complex structure of DAS data.

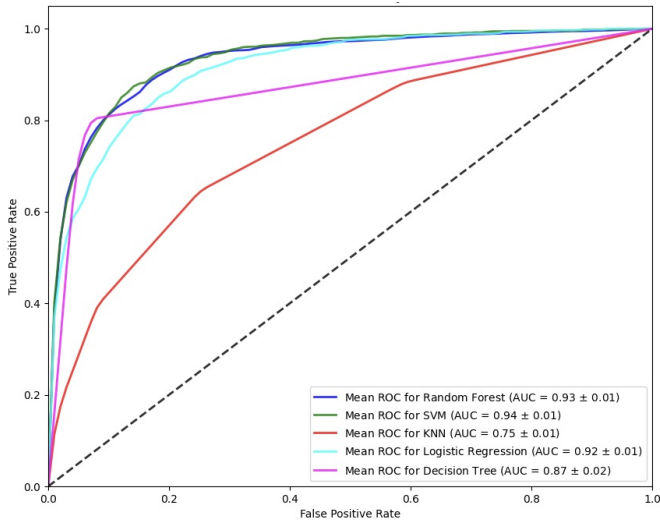


Fig. 11: Multi-class ROC curves by classifier

This ROC curve analysis emphasises the necessity of utilising multiple metrics and analyses to comprehensively assess classifier performance. While F1 scores provide a snapshot of classification accuracy at specific thresholds, ROC curves and AUC values deliver deeper insights into the classifiers' capabilities to balance sensitivity and specificity across various operational scenarios.

VII. DISCUSSION

Ethical considerations were paramount throughout the research, adhering to stringent standards. The study's design, data collection procedures, and data usage received approval from the Research Ethics Committee (REC) of Universiti Teknologi MARA (UiTM). The REC enforced a specific time frame for data collection to ensure compliance with ethical guidelines.

The research presented a novel perspective on the complex relationship between linguistic expression and mental health indicators within the Malay-speaking community on social media platforms. Analyses such as the bi-gram network and t-SNE visualisation delved into the subtleties of language usage in the context of DAS, offering complementary insights that enhance the understanding of how these conditions are linguistically manifested and discussed.

The bi-gram network analysis uncovered distinct yet interconnected linguistic patterns associated with each mental health condition, indicating both unique and overlapping language use across conditions. This overlap suggests a shared experiential basis for these conditions, despite their distinct linguistic markers. The t-SNE visualisation supported these findings by illustrating complex linguistic relationships through overlapping clusters for each condition.

Furthermore, the ROC curve analysis provided insights into the performance of various classifiers in detecting DAS conditions. The AUC values highlighted the capabilities of each classifier, with the SVM demonstrating superior discrimination

between DAS and No DAS conditions. These results underscore the potential of machine learning models to enhance the precision and sensitivity of mental health assessments using NLP.

These findings have practical implications for the development of NLP tools aimed at early detection and intervention in mental health issues, particularly in a multilingual and multicultural context.

However, the study's reliance on a single social media platform may limit the generalisability of the findings. Additionally, the imbalanced nature of the dataset and its relatively small size pose challenges to the study's broader applicability and the performance of the models on more diverse datasets.

VIII. CONCLUSION

This study has made several contributions to the field of NLP and mental health research within the Malay-speaking community:

- Development and validation of a Malay language corpus specifically designed to detect mental health indicators, thus enhancing the resources available for mental health research in the Malay language.
- Application of NLP techniques such as bi-gram networks and t-SNE visualisations to analyse the linguistic nuances associated with mental health states, providing deeper insights into the language of mental health discourse.
- Evaluation of the performance of various machine learning classifiers, highlighting the effectiveness of the Decision Tree and SVM in distinguishing between DAS and non-DAS conditions.

Future directions for this research include:

- Extending the study to include data from additional social media platforms to enhance the robustness and generalisability of the findings.
- Incorporating advanced semantic analysis techniques, such as Word2Vec or BERT, to build upon the syntactic approaches used in this study for a more comprehensive understanding of mental health conditions.
- Exploring the adaptability of the developed methodologies across different linguistic and cultural contexts.

These efforts advance the academic field of Malay NLP and offer potential enhancements for mental health interventions among Malay-speaking populations.

REFERENCES

- [1] G. Thornicroft and N. Votruba, "Sustainable development goals and mental health," **Vertex (Buenos Aires, Argentina)**, vol. 29, no. 142, pp. 300-303, Apr. 2018.
- [2] Z. Ahmad, R. Maskat, and A. Mohamed, "Harnessing Natural Language Processing for Mental Health Detection in Malay Text: A Review," in **Proc. 2023 4th Int. Conf. Artif. Intell. Data Sci. (AiDAS)**, Shah Alam, Malaysia, 2023, pp. 29-35.
- [3] P. Gooding and T. Kariotis, "Ethics and Law in Research on Algorithmic and Data-Driven Technology in Mental Health Care: Scoping Review," **JMIR Mental Health**, vol. 8, Mar. 2021, Art. no. e24668. doi: 10.2196/24668.
- [4] I. Straw and C. Callison-Burch, "Artificial Intelligence in mental health and the biases of language based models," **PLoS ONE**, vol. 15, Oct. 2020, Art. no. e0240376. doi: 10.1371/journal.pone.0240376.

- [5] S. Chancellor and M. De Choudhury, "Methods in predictive techniques for mental health status on social media: A critical review," **NPJ Digit. Med.**, vol. 3, no. 1, Art. no. 43, Mar. 2020.
- [6] K. Gupta, R. Jinad, and Q. Liu, "Comparative Analysis of NLP Models for Detecting Depression on Twitter," in **Proc. 2023 Int. Conf. Commun. Comput. Artif. Intell. (CCCAI)**, Wuhan, China, 2023, pp. 23-28. doi: 10.1109/CCCAI59026.2023.00013.
- [7] G. Fu et al., "Enhancing psychological counseling with large language model: A multifaceted decision-support system for non-professionals," **arXiv preprint arXiv:2308.15192**, Aug. 2023.
- [8] M. Conway and D. O'Connor, "Social media, big data, and mental health: Current advances and ethical implications," **Curr. Opin. Psychol.**, vol. 9, pp. 77-82, Aug. 2016.
- [9] A. Benton, G. Coppersmith, and M. Dredze, "Ethical research protocols for social media health research," in **Proc. 1st ACL Workshop Ethics Natural Lang. Process.**, Valencia, Spain, 2017, pp. 94-102.
- [10] E. Ford, S. Shepherd, K. Jones, and L. Hassan, "Toward an ethical framework for the text mining of social media for health research: A systematic review," **Front. Digit. Health**, vol. 2, Art. no. 592237, Jan. 2021.
- [11] S. Chancellor, M. L. Birnbaum, E. D. Caine, V. M. Silenzio, and M. De Choudhury, "A taxonomy of ethical tensions in inferring mental health states from social media," in **Proc. Conf. Fairness, Accountability, and Transparency**, Atlanta, GA, USA, 2019, pp. 79-88.
- [12] S. Singh, R. Joshi, P. Jain, and K. Abilash, "Human AI: Ethics and broader impact for mental healthcare," in **Emotional AI and Human-AI Interactions in Social Networking**, 1st ed. London, UK: Academic Press, 2024, pp. 191-212.
- [13] J. Kite, B. C. Foley, A. C. Grunseit, and B. Freeman, "Please like me: Facebook and public health communication," **PLoS ONE**, vol. 11, no. 9, Art. no. e0162765, Sep. 2016.
- [14] Y. Ophir, R. Tikochinski, C. S. Asterhan, I. Sisso, and R. Reichart, "Deep neural networks detect suicide risk from textual Facebook posts," **Sci. Rep.**, vol. 10, Art. no. 16685, Oct. 2020.
- [15] A. I. Rahim, M. I. Ibrahim, F. N. A. Salim, and M. A. I. Ariffin, "Health information engagement factors in Malaysia: A content analysis of Facebook use by the Ministry of Health in 2016 and 2017," **Int. J. Environ. Res. Public Health**, vol. 16, no. 4, Art. no. 591, Feb. 2019.
- [16] X. Tian, P. Batterham, S. Song, X. Yao, and G. Yu, "Characterizing depression issues on Sina Weibo," **Int. J. Environ. Res. Public Health**, vol. 15, no. 4, Art. no. 764, Apr. 2018.
- [17] Y. Ding, X. Chen, Q. Fu, and S. Zhong, "A depression recognition method for college students using deep integrated support vector algorithm," **IEEE Access**, vol. 8, pp. 75616-75629, Apr. 2020.
- [18] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in Reddit social media forum," **IEEE Access**, vol. 7, pp. 44883-44893, Apr. 2019.
- [19] U. Lokala et al., "A computational approach to understand mental health from Reddit: Knowledge-aware multitask learning framework," in **Proc. Int. AAAI Conf. Web Social Media**, Atlanta, GA, USA, 2022, vol. 16, pp. 640-650.
- [20] T. M. Wijiasih, R. N. S. Amriza, and D. A. Prabowo, "The classification of anxiety, depression, and stress on Facebook users using the support vector machine," **JISA (J. Inform. Sains)**, vol. 5, no. 1, pp. 75-79, Feb. 2022.
- [21] J. C. Eichstaedt et al., "Facebook language predicts depression in medical records," **Proc. Nat. Acad. Sci.**, vol. 115, no. 44, pp. 11203-11208, Nov. 2018.
- [22] A. Priya, S. Garg, and N. P. Tigga, "Predicting anxiety, depression and stress in modern life using machine learning algorithms," **Procedia Comput. Sci.**, vol. 167, pp. 1258-1267, Dec. 2020.
- [23] R. Musa, M. A. Fadzil, and Z. A. Zaini, "Translation, validation and psychometric properties of Bahasa Malaysia version of the Depression Anxiety and Stress Scales (DASS)," **ASEAN J. Psychiatry**, vol. 8, no. 2, pp. 82-89, Nov. 2007.
- [24] M. L. McHugh, "Interrater reliability: The kappa statistic," **Biochem. Med.**, vol. 22, no. 3, pp. 276-282, Sep. 2012.