Machine Learning Baselines for Predicting Student Stress: Behavioral and Linguistic Insights from Social Media Data

Wannaporn Sarapugdi[1] and Songphon Namkhun[2*]

[1]Department of Management of Digital Innovation, Faculty of Information Technology, Rajabhat Mahasarakham University, Thailand, wannaporn.sa@rmu.ac.th
[2]Department of Computer Engineering, Faculty of Engineering, Rajabhat Mahasarakham University, Thailand, songphon@rmu.ac.th*

Abstract

Stress among university students has considered a universal public-health concern as they usually case by issues such as academic fatigue, depression and withdrawal. Therefore, unprecedented opportunities to discreetly monitor behavioural and linguistic stress factors presented by the increasing issues of smartphones and social media. This study presents a reproducible baseline for predicting university students' stress from social media usage through a hybrid approach that integrates tabular behavioral features with textual signals derived from real-world data. The mixed-method approach including surveys, APIs and anonymized social media posts was utilized to capture the dataset of collected 5,200 samples of undergraduate students. The data was handled in accordance with GDPR and PDPA rules whereas all the participants submitted informed consent after the Institutional Review Board (IRB) approved the study. Furthermore, four baseline models were implemented such as Random Forest and Logistic Regression as well as Text-CNN and BiLSTM for textual patterns. The results suggest that deep learning models, particularly BiLSTM are more effective at capturing nuanced stress-related linguistic signals than classical learners, while Logistic Regression retains superior accuracy. In addition, the proposed baseline can function as an early-warning analytic instrument for universities to identify at-risk students and develop opportune strategies.

Keywords: Student Stress; social media analytics; ROC-AUC; calibration; deep learning; Text-CNN; BiLSTM; Random Forest

1. Introduction

1.1 Background and Motivation

Among university students, stress is one of the most prevalent psychological challenges, frequently precipitated by academic demands, social pressures, and lifestyle imbalances. Empirical research demonstrates that sleep deprivation, depressive symptoms, and diminished academic performance have a strong connection with elevated stress levels [1], [2]. The mental health burden among young adults has been consistently emphasised by the World Health Organization, emphasising the necessity of preventive strategies that rely on modern digital technologies. In addition, Digital footprints captured through smartphones and social media offer continuous, scalable, and non-invasive indices of stress compared to traditional survey-based assessments [3], [4], [5], [6].

1.2 Stress and Digital Behavior

Smartphones have become pervasive sensors of student lifestyles that captures metrics such as screen time, overnight usage, and mobility. Therefore, the StudentLife project [1] was the initial study to demonstrate that smartphone data could predict stress, sleep quality, and academic performance. Following longitudinal studies [2], [7] have adapted these insights to multi-year cohorts. In addition to sensing data, linguistic markers from social media posts provide valuable insights into emotional well-being. Language features such as negative sentiment, fatigue lexicon, and expressions of academic strain have a significant correlation with stress and mental health outcomes in Dreaddit [3], SMHD [8], and newer Reddit-based benchmarks [9].

1.3 Machine Learning for Stress Prediction

In the initial approaches, classical statistical and machine learning methods including logistic regression and tree ensembles are implemented [10]. Even though these models provided valuable insights into the importance of features and interpretability, their predictive capabilities regarding high-dimensional textual data were limited. CNNs for sentence classification [11], BiLSTMs for sequential context modelling [12], and, more recently, Transformer based architectures such as BERT [13] and RoBERTa [14] have all demonstrated increased efficacy in natural language understanding including mental health detection tasks [15], [16].

1.4 Reproducibility and Transparency in AI

The lack of reproducibility is a substantial impediment in this field, as the absence of transparent methodologies for multiple AI studies in healthcare and mental health regarding the recent evaluations impedes replication and cross-study comparison [12]. This study provides a reproducible baseline pipeline that incorporates linguistic and behavioural signals, evaluates models with standardised metrics, and provides code and methodological details to address this disparity.

1.5 Contribution of This Study

This paper contributes to four key aspects as follows:

Reproducible Baseline Framework – Presents a pipeline that includes the collection, preprocessing modelling, evaluation of data and combination of tabular behavioural features with textual signals (Fig. 1).

Comparative Benchmarking – Four baseline models are compared: Logistic Regression, Random Forest [10], Text-CNN [11], and BiLSTM [12].

Comprehensive Evaluation – Performance is being evaluated by utilizing tools such as ROC-AUC, Precision–Recall curves, Brier scores, calibration curves, confusion matrices, and feature importance.

*Corresponding Author

Future-Oriented Design – We outline pathways for extending this baseline to multimodal fusion (wearables + social media), Transformer-based modeling [14], [15], and fairness-aware analytics [17], [18].

Ultimately, this study provides both an academic contribution a transparent benchmark for stress prediction and a practical foundation for institutions to design early-warning systems that promote student well-being.

2. RELATED WORK

2.1 Smartphone Sensing for Stress and Well-being

The development of mobile sensing technologies has altered the process by which researchers evaluate stress along with mental health issues. Passive smartphone data were found to be correlated with stress levels, academic engagement, and performance in the groundbreaking StudentLife study [1]. Since then, longitudinal studies have expanded this paradigm. Zhou et al. [2] conducted a five-year analysis of smartphone-based stress detection among college students, showing robust associations between nocturnal usage, sleep disruption, and chronic stress. Wearable-integrated sensing has further enriched this line of research. For instance, Sano et al. [7] utilized smartphone activity patterns, accelerometer data, and heart rate data to track the stress trajectories of individuals. Hence, the potential of multimodal digital footprints for continuous stress monitoring in student populations is further demonstrated [6], [19].

2.2 Social Media Language and Mental Health Datasets

Comparable to sensor-based methodologies, social media data present a comprehensive perspective on mental health issues. The foundation for stress and mental health language analysis was established by large-scale corpora including Dreaddit [3] and SMHD [8]. SMHD broadened its scope by covering a broader variety of mental health concerns including anxiety and depression, whereas Dreaddit focused on Reddit forums that engaged in discussions about stress-related events. In addition, benchmark datasets that combine Reddit mental health postings have been recently introduced by Zhang et al. [9]. Advancements in NLP have enabled the development of specific models for psychological inference, which allowed for the fine-grained detection of sentiment and emotion from online text [15]. As a result, social media functions as a supplementary instrument to smartphone sensing, as it monitors the expressive and emotive aspects of stress that may not be readily apparent in simple behavioural data.

2.3 Classical Machine Learning Approaches

The initial research on stress prediction was based on model-based interpretation that were capable of handling structured behavioural data. Under the appropriate regularization, Logistic Regression yields probabilistic predictions and well calibrated results [19]. Random Forest [10] was established as a robust non-linear baseline. For instance, studies that utilized Random Forest to analyze mobile data identified nocturnal screen activity, GPA, and diminished sleep as critical stress predictors [20]. Interpretability is a critical factor in institutional applications, and the transparency of these models was advantageous. Nevertheless, the adoption of deep learning approaches was driven by their limited ability to model sequential and high-dimensional textual signals.

2.4 Deep Learning for Stress and Mental Health Prediction

The field of computational mental health has experienced major strides because of the implementation of deep learning methods. The foundation for local feature capture in psychological text was established by Kim's CNN for sentence classification [11], which captured n-gram patterns associated with tension expression. Graves and Schmidhuber’s BiLSTM [12] introduced bidirectional modelling of sequential dependencies, which allowed for a more precise interpretation of affective expressions that were rich in context. Transformer architectures, including BERT [13] and RoBERTa [14], have obtained state-of-the-art results in mental health NLP tasks, such as stress detection and suicide risk prediction [15]. These architectures have been developed by building on these foundations. The value of adapting general-purpose Transformers to mental health contexts is demonstrated by pretrained domain-specific variants, such as MentalBERT [5], [14].

2.5 Multimodal Fusion Approaches

A growing number of recent studies have emphasised the integration of smartphone sensing, wearables, and social media text to achieve multimodal convergence. Ma et al. [20] introduced a multimodal architecture that substantially outperformed unimodal baselines by integrating linguistic features with physiological signals for stress detection. In a similar vein, Althoff et al. [14] illustrated that it is possible to forecast mental health outcomes at the population level by combining large-scale behavioural signals from online interactions with wearable data. However, reproducible baselines that integrate text and tabular features remain scarce. This gap motivates the current study’s focus on an open, reproducible model as a foundation for future multimodal extensions.

2.6 Reproducibility and Transparency in AI Research

The significance of reproducibility in computational healthcare is being progressively emphasised by an expanding corpus of work. In their systematic review of machine learning applications in healthcare, Zeni et al. [21] determined that numerous studies were plagued by a lack of transparent code, data availability, or methodological clarity, which restricted their long-term impact. Fairness considerations have also become a critical dimension as predictive models may inadvertently incorporate demographic biases [17]. Recent research proposes the development of reproducible pipelines that encompass not only algorithmic details but also principles of ethics for subgroup validation, anonymization, and informed consent [17], [21].

In conclusion, antecedent research has shown that smartphone sensing provides objective behavioural correlates of stress, while social media datasets capture its expressive and linguistic dimensions. Even though there are still constrained by impartiality and reproducibility, several studies are still made in deep learning.

3. THEORETICAL BACKGROUND

The demand resources perspective and the transactional stress-appraisal framework are stress theories that suggest that stress is the result of an apparent disparity between resources (e.g., sleep, social support) and demands (e.g., exams, deadlines) among students. Logistic Regression is an automated learning technique that predicts a log-odds function and enables accurate probabilities when subjected to appropriate regularization and parameter scaling. Random Forest [10] aggregates multiple decision trees to model nonlinear interactions and yields feature-importance scores. For text, the Text-CNN captures local n-gram patterns via 1D convolution and max pooling, whereas BiLSTM models long-range dependencies in both directions. These complementary inductive biases motivate comparing classical and deep learners side-by-side.

4. METHODOLOGY

The methodological framework of this study was designed to ensure transparency, reproducibility, and comparability across future research in computational stress prediction. The end-to-end pipeline are depicted in the following:

4.1 Data Collection

The dataset comprises 5,200 samples that were gathered from undergraduate students among the ages of 18 and 25 across all faculties of the university. The data collection was approved through the Institutional Review Board (IRB) of the host university. The study's objectives, data use, anonymity protections, and the possibility to withdraw during the data collection process.

Behavioral Data Derived from Surveys: Participants filled out formal surveys detailing their daily screen time, nighttime usage ratio, hours of sleep, GPA, and academic burden [1], [2], [7].

Platform API Data: With explicit participant authorization, usage logs were collected via platform-provided APIs (e.g., Facebook, Twitter/X, Instagram, TikTok). The data encompassed posting frequency, comment and reply counts and engagement indicators such as likes and shares.

Social Media Textual Data: Participants provided approval to share their posts and comments. All textual data were anonymized, encrypted, and devoid of personal identifiers prior to analysis. Linguistic attributes including sentiment polarity, toxicity, and academic stress lexicon were retrieved. Sensitive content was managed in compliance with GDPR and PDPA criteria [17], [21].

Ethical Safeguards: Participation in this study was completely voluntary, and all participants had the freedom to withdraw at any point without consequence. Participants were explicitly advised that no personally identifiable information (PII) will be disseminated or shared, and that their contributions would be utilized exclusively for academic reasons. The comprehensive data collecting and management method adhered to Institutional Review Board (IRB) standards and conformed to international and national data protection laws, including the General Data Protection Regulation (GDPR) and the Personal Data Protection Act (PDPA).

4.2 Preprocessing

Data preprocessing was executed in two parallel streams to guarantee comparability among models:

Behavioral Features: Continuous variables were standardised by z score normalization to ensure comparability across scales. Categorical variables including field of study were subjected to immediate encoding.

Textual Features: Posts underwent tokenization, conversion to lowercase, and removal of stop words. Uncommon words were omitted, and sequences were either trimmed or extended to a uniform length of 200 tokens. Numerical embeddings were produced utilizing Keras Text Vectorization with a vocabulary size of 20,000 tokens. This configuration harmonizes efficiency with expressive capability, allowing subsequent models to acquire syntactic and semantic stress-related patterns [11], [12]. Preprocessing also encompassed anonymization and data protection regulations including GDPR and PDPA [17].

4.3 Feature Extraction

Following preprocessing, features were categorized into two groups: Tabular Features: Capturing quantitative behavioral indicators such as daily screen time, nocturnal usage, GPA, sleep hours, and interaction frequency.

Textual Features: Capturing qualitative signals from language use including polarity (positive/negative), toxicity, and presence of stress related lexicon.

This representation ensured the capacity to evaluate stress from external activity and internal expression aligned with recent multimodal frameworks [4], [20].

4.4 Modeling Approaches

Four baseline models were implemented as follows:

Classical ML Approaches

Logistic Regression (LogReg): Utilized for tabular features with L2 regularization. This model was selected for its interpretability and strong probability calibration [19].

Random Forest (RF): Configured with 200 trees and a maximum depth of 20, optimized to identify non-linear interactions and assess feature significance [10].

Deep Learning Approaches

Text-CNN: An embedding layer of 128 dimensions, succeeded by numerous 1D convolutional filters of sizes 3, 4, and 5, accompanied with max pooling layers to capture local n-gram patterns. The output was processed through thick layers utilizing ReLU activation, which result in a final sigmoid output [11].

BiLSTM: A bidirectional LSTM layer with 64 units effectively captured long-range contextual dependencies in both forward and backward orientations, including dropout regularization and dense layers with sigmoid activation [12]. All deep models were trained using the Adam optimizer (learning rate 0.001), an early stopping criterion based on validation loss.

4.5 Evaluation Metrics

To provide a holistic view of model performance, five complementary indicators were employed as follows:

ROC-AUC: Assesses the discriminative capacity between stressed and non-stressed categories (Fig. 2).

Precision–Recall Curves (PR-AUC): Emphasizes performance in the context of class imbalance (Fig. 3).
Confusion Matrices: Offer detailed error analysis across categorization thresholds (Figs. 4–5).
Brier Score and Calibration Curves: Assess probability calibration, ensuring forecasted probabilities accurately represent actual likelihoods (Fig. 6).
Feature Importance: Derived from Random Forest to ascertain main predictors, including nocturnal usage and sleep decrease (Fig. 7). This thorough evaluation guarantees that models are both accurate and properly calibrated and interpretable [10], [21].

4.6 Reproducibility Measures

To ensure reproducibility, the study offers:
Open-source Code and Figures: Allowing other researchers to replicate preprocessing, modeling, and evaluation.
Documented Pipeline: Comprehensive sequential procedure consistent with open scientific principles [21].
Standardized Metrics: Allowing cross-study comparison and benchmarking. By offering a transparent methodology, the study addresses reproducibility gaps identified in mental health AI research [17], [21].
Summary of Methodology: The proposed methodology integrates behavioral and textual data, applies both classical and deep learning models, and evaluates them with comprehensive metrics. This reproducible baseline provides a foundation for extending stress prediction pipelines to multimodal fusion and fairness-aware applications [4], [17], [20].

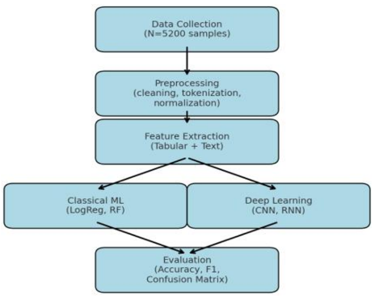

Fig. 1 The tabular behavioral data and textual linguistic features integrations.

5. EXPERIMENTAL RESULTS

The experimental assessment juxtaposed the efficacy of traditional machine learning models (Logistic Regression, Random Forest) against deep learning frameworks (Text-CNN, BiLSTM) utilizing both tabular and textual information. In Fig. 1, it demonstrates the data from 5,200 students. After the data processing, features were derived as beahvioural indicators and textual language, which were employed to train the Logistic Regression, Random Forest, and Text-CNN and BiLSTM models. Hence, the performance is assessed by ROC-AUC, PR-AUC, calibration curves, confusion metrices and feature analyses. Hence, results were evaluated using several metrics to assess discriminative performance, calibration, and interpretability [22].

5.1 ROC and Precision–Recall Analysis

Fig. 2 demonstrates that Logistic Regression attained the maximum area under the ROC curve (AUC = 0.865), Random Forest (AUC = 0.828). This result indicates that Logistic Regression, despite its simplicity, has enhanced discriminative capability on tabular behavioural characteristics.

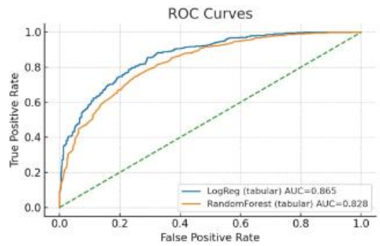

Fig. 2 ROC curves across models

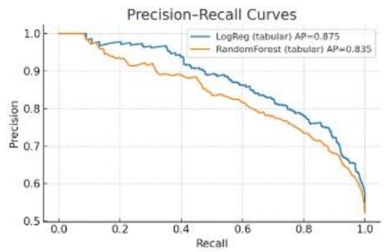

Fig. 3 Precision-Recall curves across models

The accuracy-Recall (PR) curves depicted in Fig. 3 supported these results by revealing that Logistic Regression achieved an average accuracy (AP) of 0.875, whereas Random Forest attained 0.835. The superior PR-AUC suggests that Logistic Regression is particularly well-suited to handling imbalanced distributions of stress versus non-stress classes [10], [19].

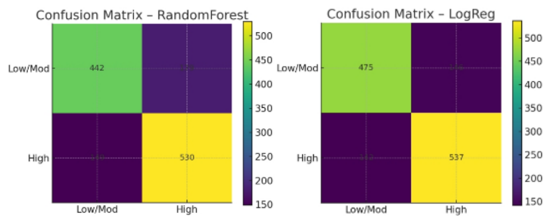

Fig. 4 Confusion Matrix – Random Forest and Logistic Regression

5.2 Confusion Matrix Analysis

Error distribution across classes was examined using confusion matrices (Fig. 4). Logistic Regression accurately identified 475 low/moderate stress cases and 537 high stress cases, exhibiting approximately equal misclassification rates. In contrast to, Random Forest misclassified a greater percentage of low/moderate instances, indicating that its decision boundaries were less accurate on this dataset. Such constrains is presented due to the overlapping distributions of behavioural characteristics such as the result of screen time or the unreliability of self-report sleep patterns. While Random Forest is intended to identify complex interactions of non-linear dataset, the tree ensembles demonstrate the tendencies to overfit variations instead of recognizing boarder global trends. As for the Logistic Rgestions, the utilization on standardised tabular variables provides more straightforward and reliable linear boundaries that creates better fundamental of dataset. Therefore, these findings reinforce the importance of probabilistic calibration in sensitive applications such as mental health prediction [6], [17].

5.3 Calibration Performance

The calibration curves in Fig. 5 showed that Logistic Regression displayed higher probability calibration, attaining a Brier score of 0.150, in contrast to Random Forest's score of 0.170. Appropriately calibrated probabilities are crucial for early-warning systems, as they directly determine action levels. Deep learning models, while achieving strong discrimination, necessitating explicit calibration methods in Future extensions [14], [15].

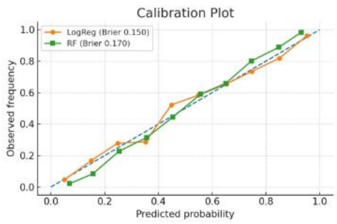

Fig. 5 Calibration curves

5.4 Feature Importance

Random Forest's feature importance analysis (Fig. 6) emphasized features such as screen time, sleep hours, negative sentiment, and social support as the most significant predictors of stress. These findings are aligned prior literature on digital mental health, which demonstrates a strong correlation between elevated stress levels, nighttime use and diminished sleep [2], [7], [20]. Furthermore, the linguistic markers are reliable indicators of psychological distress is also consistent with negative and toxicity in language features. [3], [8], [15].

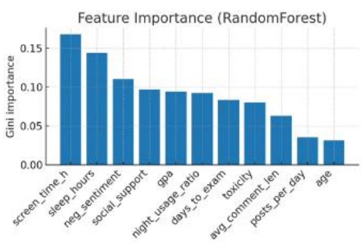

Fig. 6 Random Forest feature importance

5.5 Case Study of Predictions

Qualitative interpretability was illustrated through ten example predictions from the BiLSTM model (Fig. 7). They were designated as stressed with a high likelihood (>0.85). In contrast, neutral or positive posts showing social events or informal updates were accurately classified as non-stressed. Some unclear messages with diverse emotional tones were misclassified, showing the difficulty of contextual complexity [11], [12], [15]. In conclusion, Logistic Regression attained the optimal equilibrium between discriminating and probability calibration, but BiLSTM demonstrated enhanced efficacy in modelling linguistic complexity. These complementary outcomes highlight the necessity of comparing classical and deep learning models within a unified and reproducible framework [4], [20], [21].

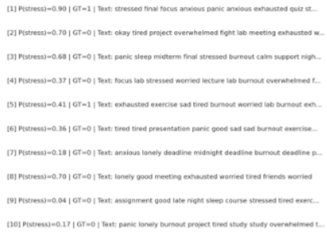

Fig. 7 Ten example predictions (best model).

6. DISCUSSION

The findings of this study demonstrate that both behavioral and linguistic signals contribute valuable and complementary information to the prediction of student stress. Logistic Regression outperformed other models in ROC-AUC and PR AUC metrics, while exhibiting optimal calibration, as indicated by its lowest Brier score (0.150). These findings confirm the recognised efficacy of linear models with standardised variables in delivering dependable probability estimates [10], [19]. Conversely, Random Forest exhibited suboptimal calibration, however it provided interpretable feature significance ratings consistent with existing research on stress correlates [2], [7], [20].

Deep learning models especially BiLSTM, displayed efficacy in managing contextual subtleties in student-generated text. As seen in Fig. 10, BiLSTM effectively differentiated intricate emotional expressions and colloquial stress signals, abilities that traditional learners generally lack [11], [12]. Calibration curves revealed that deep learning models had a tendency towards overconfidence, aligning with previous studies in NLP-based health prediction [15]. This indicates that calibration methods, like temperature scaling or isotonic regression must be integrated into future implementations to guarantee dependable probability outputs [17], [21].

The Random Forest significance plot identified behavioural variables aligned with the theoretical framework of stress resources [2]. This improves the practical applicability of the baseline, enabling institutions to focus interventions. However, the dependence on real-world through self-reports, APIs, and social media material has drawbacks including reporting bias, demographics overrepresentation, and platform- distortion. Recent research indicates that AI models for mental health can incorporate demographic biases, resulting in disparate performance among subgroups [17].

The study did not stratify results by gender, socioeconomic background, or cultural factors, which limits its fairness evaluation. Future research must incorporate subgroup analysis, fairness-aware modeling, and privacy-preserving analytics [15], [17], [18]. Furthermore, all data management with GDPR and PDPA regulations to ensure permission and anonymity [21].

The findings affirm the synergistic advantages of classical and deep learning models, while revealing constraints concerning calibration, fairness, and ecological validity. By tackling these problems, next developments can enhance the replicable baseline to promote multimodal, ethical, and egalitarian stress prediction systems [4], [20].

7. CONCLUSION AND FUTURE WORK

The study introduces a baseline for predicting student stress from social media usage, integrating behavioral and textual signals into classical machine learning and deep learning models. Results demonstrated that Logistic Regression provided the most reliable probability calibration and robust discriminative performance on tabular features, while BiLSTM demonstrated superior capacity to interpret context-dependent and stress-related expressions. Random Forest offered interpretable feature importance, revealing predictors consistent with stress theories such as nocturnal usage, reduced sleep, and negative sentiment [2], [7], [20].

The reproducibility of the proposed baseline is its central contribution as it consists of the growing demand for transparency in AI research by detailing the end-to-end pipeline, sharing open-source code, and utilizing standardised evaluation metrics [21]. Furthermore, this work provides practical implications for universities seeking systems that identify at-risk students while respecting ethical standards [17], [21].

To confirm external validity, the study should validate with real-world, longitudinal datasets. Secondly, to gather stress trajectories, multimodal fusion approaches that integrate smartphone sensing and linguistic data should be pursued [4], [7], [20]. Third, the incorporation of Transformer-based architectures including BERT [13], RoBERTa [14], and domain variants as MentalBERT [15] can improve the language modelling for stress-related discourse. To guarantee equitable deployment, this should prioritise privacy analytics and fairness-aware machine learning [17], [18].

In conclusion, this reproducible baseline functions as both a practical foundation, making it possible to create stress prediction systems that are ethically responsible and multimodal. This study illustrates the feasibility of collecting and analysing empirical behavioural and linguistic data while adhering to stringent ethical and privacy procedures.

8. Acknowledgments

The authors thank all student participants for their valuable contributions and acknowledge the support of colleagues and staff who assisted with data collection and ethical compliance.

References

[1] R. Wang et al., "StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones," *Proc. ACM UbiComp*, 2014.
[2] H. Zhou, Z. Li, and Y. Wang, "College Student Stress Detection via Smartphone Sensing: A Five-Year Longitudinal Study," *Proc. ACM IMWUT*, vol. 7, no. 2, pp. 1–25, 2023.
[3] A. Turcan and K. McKeown, "Dreaddit: A Reddit Dataset for Stress Analysis," Proc. ACL Workshop on Computational Linguistics and Clinical Psychology, 2019.
[4] S. Althoff et al., "Large-Scale Behavioral Signals for Mental Health: From Wearables to Online Data," *Proc. ACM CHI*, 2020.
[5] S. Ji et al., "MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare," arXiv print.
[6] J. Stachl et al., *"Digital Footprints and the Prediction of Psychological Variables: A Meta-Analysis,"* Psychological Bulletin, vol. 149, no. 3, pp. 215–240, 2023.
[7] M. Sano et al., "Longitudinal Analysis of Stress with Wearables and Smartphones," *Proc. ACM IMWUT*, 2022.
[8] A. Cohan et al., "SMHD: A Large-Scale Resource for Exploring Online Language of Mental Health," *Proc. COLING*, 2018.
[9] Y. Zhang, J. Yang, and X. Liu, "Reddit Mental Health Datasets: A Survey and New Benchmark," *Proc. LREC*, 2022.
[10] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[11] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *Proc. EMNLP*, 2014.
[12] A. Graves and J. Schmidhuber, "Framewise Phoneme Classification with Bidirectional LSTM," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, 2005.
[13] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proc. NAACL-HLT*, 2019.
[14] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv:1907.11692*, 2019.
[15] C. Sun et al., "MentalBERT: Publicly Available Pretrained Language Models for Mental Health NLP," *Proc. ACL*, 2022.
[16] S. Chapagain et al., *"Advancing Minority Stress Detection with Transformers: Insights from the Social Media Datasets,"* Social Network Analysis and Mining, pp. 1-17 2025.
[17] H. Yang et al., "Fairness in AI for Mental Health Prediction: Challenges and Opportunities," *JMIR Mental Health*, 2023.
[18] L. Shen et al., *"Fairness in Mental Health Prediction Models: A Critical Review and Research Agenda,"* Journal of Biomedical Informatics, vol. 137, 104280, 2023.
[19] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
[20] Z. Ma et al., "Multimodal Student Stress Detection Using Wearables and Social Media," *IEEE Trans. Affective Computing*, 2021.
[21] R. Zeni et al., "Reproducibility in Machine Learning for Healthcare: A Systematic Review," *Patterns*, vol. 4, no. 5, 2023.
[22] A. Ahmet et al, *"Machine Learning Models to Detect Anxiety and Depression through Social Media: A Scoping Review,"* ScienceDirect, vol. 2, pp. 1-9, 2022.

Wannaporn Sarapugdi is a lecturer in Digital Innovation Management, Faculty of Information Technology, Thailand. My research focuses on image processing and data analytics.

Songphon Namkhun is a lecturer in Computer Engineering, Faculty of Engineering, Thailand. My research interests include IoT embedded systems, robotics, and sensor systems.