|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Authors | Title | Algorithms Used | Models Used | Dataset | Accuracy | Findings | Limitations |
| Vandana, et al (2023) | A hybrid model for depression detection using deep learning | CNN, LSTM, Bi-LSTM | Textual CNN, Audio CNN, Hybrid LSTM/Bi-LSTM | DAIC-WOZ Database (189 sessions) | Textual CNN: 92 Audio CNN: 98 Bi-LSTM: 88% | Audio > Text (98% vs 92%). Bi-LSTM > LSTM. Hybrids = more robust. | Imbalanced data (4:1), Bi-LSTM slow (5+ hrs), limited to DAIC-WOZ. |
| Lamia Bendebane et al (2023) | *A Multi-Class Deep Learning Approach for Early Detection of Depressive and Anxiety Disorders Using Twitter Data* | Deep Learning (CNN, RNN, LSTM, GRU, BiRNN, BiLSTM, BiGRU), Grid Search | Hybrid models (e.g., CNN-BiGRU, CNN-BiLSTM) | 3.17M tweets (English) | **93.38** (CNN-BiGRU) | Multi-class > Binary Detects depression vs. anxiety well Grid search tuned learning rate | Labeling issues Not tested on non-English tweets Needs clinical validation |
| Nikhil Goel et al (2024) | *Automated Depression Detection System: Integrating Sentiment Analysis and Behavioural Data* |  | Hybrid (SVM + Decision Trees), Neural Networks | 1,000 subjects (text + wearable data) | **90** (Hybrid Model) | Sentiment (85%) + behavior (r=0.7) Hybrid model: F1 = 0.89 30% symptom reduction post-intervention | Reliance on self-reported data (bias risk). Contextual ambiguity in sentiment analysis. Device variability affects behavioural data quality. |
| Shumaila Aleem et al (2022) | *Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions* | SVM, RF, KNN, DT, AdaBoost, CNN, LSTM, DCNN, XGBoost | Classification, Deep Learning, Ensemble | EEG, social media (Twitter, Reddit), clinical records (PHQ-9, BDI-II) | 76.6–98.32 | SVM and RF are robust; EEG-based DL models achieve high accuracy; multimodal approaches show promise. | Small sample sizes, lack of standardized datasets, limited clinical applicability. |
| Mumtaz & Qayyum (2019) | *EEG-based DL model for diagnosing unipolar depression* | 1DCNN, LSTM | Deep Learning | EEG (30 healthy, 33 MDD) | 98.32 | High accuracy in EEG-based depression classification | High accuracy in EEG-based depression classification |
| Faye Beatriz Turnaliuan et al (2024) | *Development of a two-stage depression symptom detection model: application of neural networks to twitter data* | LSTM with Dropout, GRU, CNN, Naïve Bayes, Random Forest | Two-stage model (Binary + multi-class) | 86,163 tweets (English/Filipino) annotated with 13 depression categories | Stage 1: 91 (F1: 0.90)  Stage 2: 83 (F1: 0.81) | Two-stage model: Binary detection + 6 symptom types LSTM + Dropout = best performance | Errors from word associations, negation, imbalance Limited to English/Filipino; excludes regional languages |
| Rafael Salas-Zárate (2022) | Detecting Depression Signs on Social Media: A Systematic Literature Review | SVM, Logistic Regression, Neural Networks, Random Forests | Word Embedding, N-grams, Bag of Words, Tokenization | Twitter, Reddit, Facebook, Instagram, Weibo, NHANES | N/A | Twitter + SVM/embeddings most used Python tools, cross-validation standard | Limited to studies from 2016-2021. Focused mainly on English-language platforms. |
| Arora and Arora (2019) | Mining Twitter Data for Depression Detection | SVM, Decision Trees | N-grams, Bag of Words, Stemming | Twitter (3754 tweets) | N/A | Compared SVM and Decision Trees for depression detection. Found SVM to be more effective. | Small dataset. Limited to Twitter. |
| Nadeem (2016) | Identifying Depression on Twitter | SVM, Neural Networks | Bag of Words, TF-IDF | Twitter (1,253,594 tweets) | N/A | Used TF-IDF and Bag of Words for feature extraction. Compared SVM and Neural Networks. | Large dataset but limited to Twitter. No accuracy reported. |
| Yazdavar (2020) | Multimodal Mental Health Analysis in Social Media | SVM, Neural Networks | Word Embedding, LIWC, Cohen's Kappa | Twitter (8770 users) | N/A | Combined linguistic and behavioural features for depression detection. Used Cohen's Kappa for validation. | Complex multimodal approach may not be scalable. |
| Chiong (2021) | A Textual-Based Featuring Approach for Depression Detection Using Machine Learning Classifiers and Social Media Texts | SVM, Neural Networks | N-grams, Bag of Words | Twitter, Facebook (22,191 records) | N/A | Compared SVM and Neural Networks for textual analysis. Found SVM to perform better. | Limited to textual features. No accuracy metrics. |
| Katchapakirin (2018) | Facebook Social Media for Depression Detection in the Thai Community | SVM | LIWC, RapidMiner | Facebook (35 users) | N/A | Developed a depression detection algorithm for Thai-language Facebook posts. Used LIWC for feature extraction. | Small dataset. Limited to Thai language |
| Wongkoblap (2019) | Predicting Social Network Users with Depression from Simulated Temporal Data | Neural Networks | Word Embedding, Softmax Function | Simulated data | N/A | Used temporal data and word embedding for depression prediction. Applied Softmax for classification. | Simulated data may not reflect real-world scenarios. |
| Bazen Gashaw Teferra (2024) | Screening for Depression Using Natural Language Processing: Literature Review | SVM, Logistic Regression, Neural Networks, Transformers (BERT, GPT) | Sentiment Analysis, Linguistic Markers, Word Embeddings, LLMs | DAIC-WOZ, Weibo, Twitter, Reddit | 82.3 - 91 | NLP (sentiment, LLMs) = high accuracy Key issues: ethics, cultural sensitivity | Limited databases, no meta-analysis English/Chinese focus limits generalizability |
| Rathners (2017) | How did you like 2017? Detection of language markers of depression and narcissism | Logistic Regression | LIWC-based features | Personal narratives (220 participants) | R² = 0.104 | Demonstrated the use of LIWC for detecting linguistic markers of depression in personal narratives. | Small dataset; limited to specific narrative context. |
| Prabhu (2022) | Harnessing emotions for depression detection | LSTM | Word2vec | DAIC-WOZ (189 sessions) | 82.3 | High accuracy: LSTM + Word2Vec for emotion-based detection | Clinical data only; may not generalize to social media |
| Islam (2018) | Depression detection from social network data using machine learning techniques | Decision Tree | LIWC | Facebook comments (7145 comments) | F-measure = 0.71 | Used decision trees with LIWC features to detect depression in Facebook comments. | Focused on Facebook; may not apply to other platforms. |
| Choudhury (2021) | Predicting depression via social media | SVM | LIWC (22 linguistic styles) | Twitter (554 users) | 72.4 | Identified linguistic styles associated with depression on Twitter using SVM. | Limited to Twitter; potential bias in user selection. |
| Stankevic (2018) | Feature engineering for depression detection in social media | SVM | Word Embeddings | CLEF and eRisk 2017 (887 users) | F₁-score = 63.4 | Word embeddings + SVM used; feature engineering was key | Moderate performance; dataset limitations. |
| Lopez-Otero (2017) | Depression detection using automatic transcriptions of de-identified speech | SVM | GloVe | DAIC-WOZ (189 sessions) | F₁-score = 73 | Applied GloVe embeddings with SVM for speech-based depression detection. | Small dataset; limited to clinical settings. |
| Mallol-Ragolta (2019) | A hierarchical attention network-based approach for depression detection | Hierarchical Attention Network | GloVe | DAIC-WOZ (189 sessions) | UAR = 0.66 | Proposed a hierarchical attention network for depression detection using GloVe embeddings | Complex model; requires large datasets for training |
| Dinkel (2020) | Text-based depression detection on sparse data | SVM | ELMo | DAIC-WOZ (189 sessions) | F₁-score = 84 | Achieved high performance using ELMo embeddings for sparse data depression detection. | Limited to specific datasets; may not generalize. |
| Rutowski (2022) | Depression and anxiety prediction using deep language models and transfer learning | Transformers | GloVe | American English spontaneous speech (16,000 sessions) | AUC = 0.8 | Transformers used for prediction; transfer learning proved effective | Focused on English speech; may not apply to text data. |
| Korti (2022) | Depression detection from Twitter posts using NLP and machine learning techniques | LSTM | Word Embeddings | Twitter | 91 | Achieved high accuracy with LSTM for Twitter-based depression detection. | Limited to Twitter; potential bias in data collection. |
| Tejaswini (2024) | Depression detection from social media text analysis using hybrid deep learning | FastText + LSTM | FastText | Reddit and Twitter (13,000 posts) | 87 | Combined FastText and LSTM for high-accuracy depression detection on social media. | Focused on English platforms; may not generalize to other languages. |
| Senn (2022) | Ensembles of BERT for depression classification | BERT | Transformers | DAIC-WOZ (189 sessions) | F₁-score = 0.62 | BERT ensembles used for depression classification in clinical interviews | Small dataset; computational complexity. |
| Hayati (2022) | Depression detection on Malay dialects using GPT-3 | GPT-3 | Few-shot Learning | Interview questions (53 participants) | F₁-score = 0.64 | Applied GPT-3 for few-shot learning in Malay dialect depression detection. | Small dataset; limited to specific cultural context. |
| Németh (2022) | Bio, psycho, or social: supervised machine learning to classify discursive framing | DistilBERT | Transformers | SentiOne (80,000 posts | 73 | Used DistilBERT to classify discursive framing of depression in online health communities | Focused on discursive framing; not direct depression detection. |