International Research Journal Multidiciplinary Sciences ISSN Awaited
Vol 1 Issue 3 March 2025 PP 9-27

Using Machine Learning For Sentiments And Behavioral Analysis In Digital Communication To Provide Mental Health Insight And Support

[1]Prof.QuisarAlam(quisarquaisar.alam@niu.edu.in)
[2]Prof.Yazdani Hasan(yazhassid@gmail.com)
[3]Mr.Shahid
[1,2,3]Noida International University

Abstract

The goal of the review is to describe how machine learning has demonstrated exponential progress in consuming vast amounts of data and producingaccurate results on par with human intelligence. It offers a glimpse of the future, when advanced data, analysis, and analytical models work together to aid countless individuals with health problems. The current use of machine learning (ML) in the healthcare industry is reviewed in this paper, along withits limitations, predictive analysis, and areas that require further investigation and are difficult to diagnose.

New Findings: We have examined thirtypapersonthe use ofmachine learning (ML) for the detection of mentalstress. These papers used a varietyofsources, including social networking sites, blogs, discussion forums, student records,the Questioner technique, clinical datasets, real-time data (audio, video, and driving tasks), bio-signal technology (EDG, EEG), wireless devices, and suicidal thoughts. Together, these studies demonstrate the great accuracy and promise of machine learning algorithms in mentalhealth, as wellas which ML algorithm produces the best outcomes.

Summary: Many sectors, such as traditional clinical trials, which are insufficient to gather all of a person's information, have emerged with the development of machine learning. Currently, these disorders are classified under the DSM-V stage in order to identify them early on and begin treatment before anything goes wrong. It has changed the mental health practice by cutting downonstandtime, making it easier and more convenient for people to receive better healthcare whenever they need it.

Keywords: Mentalstress;Sentimentanalysis;SVM;NaïveBayesian classifier;Twitter; Depression; Machine learning

1. Introduction

Advancements intechnologyhave ledtothediscoveryanddevelopment of innovative methods with significant potential to assist patients facing complex medical conditions. As technology progresses, patients are among the first to benefit, as it enables the identification of optimal treatment options.

Globally, healthcare organizations are increasingly adopting digital and automated systems. Machine Learning (ML) has the ability to collect and analyze vast amounts of data, generating intelligent solutions for diagnosing and treating patients with mental and emotional health conditions. The rising global population has placed immense pressureon the healthcare sector to improve treatment and service delivery.

According to McKinsey, big data and ML have the potentialto generate annual revenues of $100 billion. ML has revolutionized healthcare by making groundbreaking advancements. For example, Google has developed an ML algorithm capable of detecting cancerous tumors, while Stanford University has applied deep learning techniques to identify skin cancer.

This article provides a comprehensive review of ML techniques in mental health, their applications across various healthcare domains, a literature survey, performance metrics, a summary of relevant studies, limitations, futureresearchdirections, recommendations, andchallenges in diagnosing complex health conditions.

MLinMentalHealth

The term "Machine Learning" (ML) was first introduced by Arthur Samuel in 1952. Over the years, MLapplications in medicine have grownsignificantly. It enablesmachinesto"learn"automaticallyandpredict outcomeswithout human intervention. Deep Learning (DL), a subset of ML, is widely used inhealthcare, assisting both patients and medical professionals.

ML plays a crucial role in fields such as oncology, radiology, cardiology, and pathology, where complex datasets are analyzed to identify patterns, aiding clinicians in making better decisions by interpreting medical images and reports.

According to the World Health Organization (WHO), approximately 1.94 billion people worldwide suffer from mental health disorders, with anxiety beingthemostcommoncondition,affecting248millionindividuals.

Depression is a severe mental health issue that can lead to suicidal tendencies, contributing to an estimated 8 million deaths annually (14.3%). Early recognition of stress symptoms, whether short-term or long-term, can prevent suicidal thoughts and improve mental health outcomes.

Similarly, the healthcare industry generates vast amounts of data. Many countries have started digitizing patient information through Electronic Health Record (EHR) software, which stores data such as medical records, prescriptions, and billing information. The WHO states that depression often leads to mentaldisorders, highlighting the need for intelligent systems capable of detecting early symptoms and learning from data to provide accurate and timely predictions about a person's emotional state.

MLTechniquesforBigDataAnalysis

MachineLearning(ML) isanintegralcomponent of ArtificialIntelligence(AI) that automatespredictive analytics. It isparticularlyeffective inprocessing and analyzing large datasets collected from various sources. According to IDC, the volume ofdata is expected to grow by 50 times, reaching 5.2 ZB for analytical processing by 2025. Unlike traditional methods, which focus on interpretation, ML is prediction-driven. It enables systems to learn from data, identifypatterns, and generate insights that are valuable for decision-making.

ML is broadly classified into three categories: Supervised Learning, Unsupervised Learning, and Reinforcement Learning. Additionally, Deep Learning (DL) is a specialized subset of ML.

SupervisedLearning(SL)

In this learning approach, a dataset acts as a guide, training the model to learn from observations. Once trained, the model can make predictions when new dataisintroduced.Mathematically,thisprocessinvolvesmapping inputs (X)to outputs (Y) using an algorithm: Y = f(X).

Supervised Learning is widely applied in healthcare, where it helps recognize patterns and assist clinicians in decision-making. A common example is text classification, which analyzes text-based data to detect sentiment. In mental health diagnostics, it can classify cases of Major Depressive Disorder (MDD) by categorizing posts as depressed, non-depressed, positive, negative, or neutral.

UnsupervisedLearning(USL)

Unlike Supervised Learning, Unsupervised Learning does not rely on labeled data or external supervision. It works with input variables (X) but lacks predefined output variables (Y). Instead, the algorithm identifies hidden patterns in data and organizes it into meaningful clusters.

Common techniques used in Unsupervised Learning include K-Means clustering, Hierarchical clustering, K-Nearest Neighbors (KNN), and Principal Component Analysis (PCA). In healthcare, genetic analysis uses clustering techniques to study DNA patterns, aiding in biological evolution research and cancer classification based on genetic computations.

ReinforcementLearning(RL)

Reinforcement Learning is based on a "trial-and-error" approach. The system interacts with its environment, learns from feedback, and optimizes its actions to achieve the best possible outcome.

Inhealthcare, Reinforcement Learning has promising applications, although its real-worldimplementationfaceschallenges.Somenotableexamplesincludeits use in the treatment of lung cancer, epilepsy, and bioinformatics. Byanalyzing past patient data and adapting to new information, RL-based models can assist in personalized treatment planning and predicting disease progression.

DeepLearninginHealthcare

DeepLearning(DL) hassignificantlyenhancedhealthcarebyassisting medical professionals and improving patient care. It can learn autonomously,processing unlabeled and unstructured data to extract meaningful insights.Deep Learning AI mimics the functioning of the human brain, enabling applications such as object recognition, speech processing, languagetranslation, and decision-making. In the healthcare industry, DL-basedsolutionsareusedfor chatbots,medicalimaginganalysis, cancer detection, and the identification of rare diseases.

MachineLearninginMentalHealth

In specialized healthcare fields like bioinformatics, ML has made significant progress by analyzing complex datasets. Researchers are increasingly usingML techniques for diagnosing mental health disorders. The five most commonly used ML algorithms in mental health studies include:

- SupportVectorMachines(SVM)
- RandomForest (RF)
- K-NearestNeighbors(KNN)
- GradientBoostingMachine(GBM)
- NaïveBayes(NB)

According to Cho et al. (2019), SVM, GBM, RF, and NB have been widely applied in mental health research. The primary purpose of ML techniques is to analyze large datasets and extract valuable insights for better decision-making.

The healthcare sector primarilyutilizes Supervised and Unsupervised Learning for medical data analysis. Additionally, Reinforcement Learning (RL) is also used for predictive modeling and complex data-driven decision-making in healthcare applications.

2. Background

SocialMediaandMentalHealthPrediction

Intheir 2017study,Aldarwishand Ahmad exploredtheuseofuser-generated content (UGC) fromsocial networking platforms like Facebook, Twitter, and Instagram to predict mental health conditions. Researchers believe that analyzingsocialmediapostscanhelp identifyindividualsexperiencing mood fluctuations, stress, anxiety, and loss of hope, including those who have not been formally diagnosed.

AccordingtotheWorldHealthOrganization(WHO),suicide isoneofthe leading causes of death among individuals aged 15 to 29 worldwide.

MachineLearninginMentalHealthandSentimentAnalysis

Machine Learning (ML) has widespread applications in medical diagnosis, speech recognition, image processing, and Natural Language Processing (NLP).It enablesresearchersto extractvaluableinsightsfrom largedatasets and develop intelligent systems for mental health assessment.

Diagnostic questionnaires such as the Beck Depression Inventory (BDI) and theCenterforEpidemiologicStudiesDepressionScale(CESD-R)(Hussainet al., 2015) arecommonlyused for assessing patients' mentalhealthconditions.

EmotionArtificialIntelligenceandTextAnalysis

Deshpande and Rao (2017) highlighted that Emotion Artificial Intelligence is an evolving field focused on text analysis. With the rapid expansion of digital media,bothtext andimage-baseddatasetsarenowusedfor sentiment analysis. Researchersclassifysocialmediaposts, suchas tweets, asnegative, neutral, or positive, which helps in detecting depression.

Today, people frequently express their thoughts, emotions, and opinions through social networking sites, making text-based communication one of the most widelyused formsofinteraction.In EmotionAI,textualdataisleveraged for sentiment detection using various ML techniques.

ImpactofMentalHealthDisorders

Stress is often associated with negative life experiences and is considered a formofmentaldistress. According to globalstatistics, mentalhealthdisorders

Vol 1 Issue 3 March 2025 PP 9-27

haveseverecconsequencesonthebody, leadingtoconditionssuchas panic attacks, anxiety, fear, depression, substance use disorders, schizophrenia, eating disorders, and post-traumatic stress disorder (PTSD).

Reports indicate that around 13% ofthe global population suffers from mental health and substance use disorders. While occasional feelings of sadness are normal,persistentnegativeemotionsmayindicate mentalillnessordepression, requiring timely intervention.

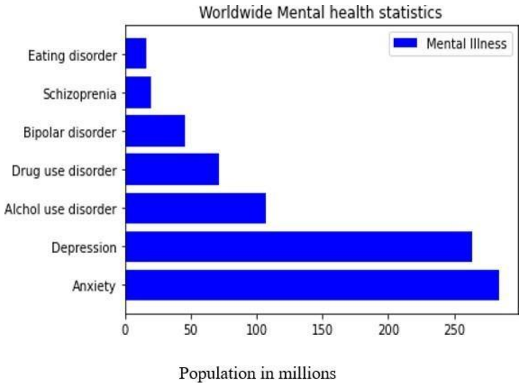

Figure.1WorldwideStatisticsofMentalHealthandSubstanceusedisorder(inmillions)

3. Methods:StudySelectionandPerformanceMeasures

ReviewofResearchonMentalIllnessDetectionUsingMachineLearning

A comprehensive review of research papers focusing on the detection ofmental illness using various Machine Learning (ML) techniques has been conducted. The research papers were sourced from PubMed, Google Scholar, ScienceDirect, conference proceedings, and academic journals. The selection process involved using relevant keywords such as mental illness diagnosis, sentiment analysis, depression detection, and machine learning.

DataSourcesforStressDetection

The studies reviewed primarily focus on stress detection using social media posts fromplatforms like Twitter and Facebook, as well as clinical records and biometric sensor data (e.g., Heart Rate Variability(HRV), Electrocardiography (ECG), and Electroencephalography (EEG)).

Table 1 presents details of 30 studies, summarizing key aspects such as study objectives, datasets used, accuracy levels, methodologies, and ML algorithms applied in the field of mental health prediction.

PopularMentalHealthHashtagsandMLAdvancements

Social media users often express their mental health struggles through widely used hashtags suchas #depression, #anxiety, and #sadness, whichare analyzed for sentiment analysis (Fig. 2). Over the years, ML has significantlycontributed to healthcare by improving diagnosis, treatment, medical data analysis, and decision-making in clinical settings.

GrowthofResearchinMentalHealthDetectionUsingML

The number of publications on mental illness detection using ML techniques has steadily increased between 2011 and 2020 (Fig. 3). Different ML algorithms have been applied to various datasets to detect mental stress, with accuracy varying based on the dataset size and sample composition.

ComparisonofMLAlgorithmPerformance

A comparative analysis (Fig. 4) highlights the accuracy levels of Support Vector Machines (SVM) and Naïve Bayes across different datasets, including:

- SocialMediaPosts(Facebook,Twitter)
- SentimentAnalysisonTwitterandFacebookPosts
- UniversityStudentRecords
- BiosensorData(EEG)

These findings demonstrate the effectiveness of Machine Learning in mental health analysis, paving the way for further advancements in stress and depression detection.

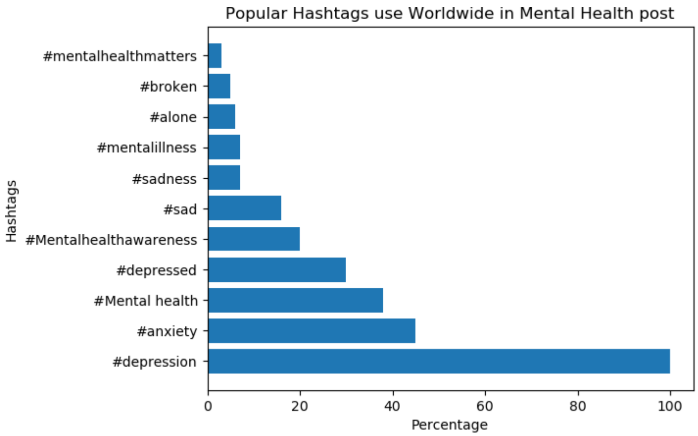

Figure.2TopHashtagsusedindepressiononInstagram,Twitter, and Facebook

International Research Journal Multidiciplinary Sciences ISSN Awaited
Vol 1 Issue 3 March 2025 PP 9-27

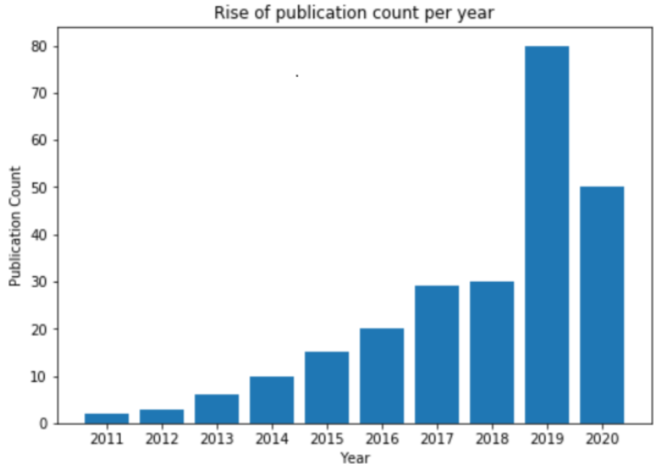

Figure.3Publicationcountincreasesfrom(2011-2020)inmentalhealthusingML

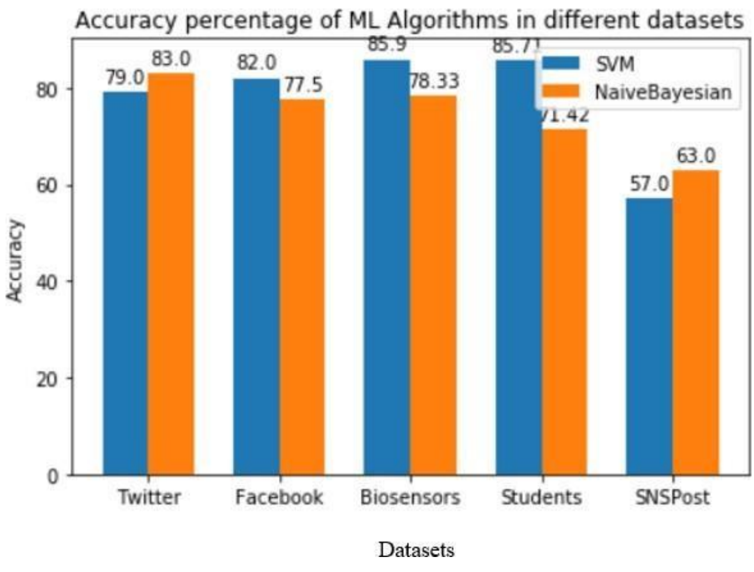

Figure.4AccuracypercentageofSVMandNaïveBayesianclassifierondifferentdataset

Table1SummaryofMLTechniquesusedtodetectMentalStress,Dataset,Methodand

Author	Method	PurposeofStudy	Dataset	ML Techniques	Accuracy
SocialNetworkSites					
Maryametal.,(2017)	Quantitative	Predicting stress from UGC-UserGenerated ContentinSocialmedia sites(Facebook,Twitter, LiveJournal)classifyon basisofmoodand negativism and BDI-questionnaire.	6773 post, 2073 depressed,4700 non-depressed post(textual)	SVM Naïve-Bayesian	57%63%
Gyeongcheolet al., (2018)	Qualitatively	AnalysisofML Algorithms helps Diagnosing mental illness, properties and theirlimitationsandhow algorithmsimplemented	Not reported	SVM, GBM,KNN, Naïve Bayesian,K Nearest, Random Forest	75% (highest accuracy reported by SVMclassifier due to itsdata sparsity)
Reshmaet al.,(2019)	Quantitative	Detecting sentiment analysisontwitterusing Tensi Strength frameworkFromsocial media text. Lexicon approachtodetectstress and relaxation	Large datasets dividedinto 100record (textual)	SVMNB WSDandn-gram	65%(Precision True Positives oversumofTP and FP) 67% (Recall True Positiveoverthe sum of TP and FN)
Jamiletal., (2015)	Quantitative	Automaticallyclassifies the MDD diagnosis from individuals Facebook profile in addition questionnaire techniqueBDIand CESD-R.	Facebookpost	SVM KNN	Sadness-53.66% Late night activity-13.33% Confused-26.66%
Mandaret al.,(2017)	Quantitative	Detectdepressionusing emotional analysis on TwitterfeedsusingNLP basedoncuratedword-list(negativeor neutral)	10,000 Tweet Using Twitter API	SVMF-1 ScoreNaïve BayesNLP	79%(linearand non- linear) 83%(text classification)
Moha.etal., (2018)	Mixed	Comparativeanalysisof differentMLtechniques and deep learning, and developahybridmodel for sevtiment analysis	1,048,588 tweets (positiveand negative classification)	Naive Bayes Random Forest Decision TreeCNN Hybrid Model	77.5% 73.8% 72.5% 79.6%(wordto vector83.6%)

Author	Method	PurposeofStudy	Dataset	ML Techniques	Accuracy
SocialNetworkSites					
Chawreetal., (2020)	Qualitative	Facebookpostsusedas token.Featureextraction from social interaction and classifyas positive and negative.	Not reported	TSVM SVMNaïve Bayes Random forest Decision Tree Adaboosted D-Tree	84% (extract post) 82% (classifypost) 77.5% 73.8% 72.5% 67%
Meghaetal., (2018)	Mixed	Analyze existing Sentiment analysis Techniques and to Improve the overall accuracy using hybrid sentimentclassification model for tweets.	1,600,000 tweet, 1000 positiveand 1000 Negative review7086 sentences	SVM Adaboosted D- Tree Decision Tree	82% 67% 84%

ClinicalDatasets					
Ravinderet al.,(2019)	Quantitative	Calculate stress of University students one week before their exams and usage ofinternet using MLtechnique	Record of 206 students(Jaypee Institute of Information Technology)	SVM Randm Forest Naïve Bayes KNNPSS	85.71% 83.33% 71.42% 55.55%
Christ.et al.,(2018)	Quantitative	Predicting persistent symptoms ofdepression in age more than 65	Data of (284 patients) based ondemographic and physcometricfor continuous 12months	LR GB	69% 74%
Hugoetal.,(2017)	Quantitative	Byevaluatethebehavior (depression, stress,self-esteem) of an adolescent and detectthe suicidal tendency. Propose a desktop tool.	Dataset of adolescent with suicidaltendency in Peru 10000 Instances800real record(suicide attempts)	JRIPalgorithm C4.5(decision tree family) Naïve Bayes	97.4% 98.4% 98.65%
Ashleyet al.,(2020)	Quantitative	Predictingmentalillnessin adolescent	7,638twinsfrom child and adolescentTwin study	Randomforest SVM	75% 75%

PhysiologicalSensors					
Chandrasekhar *etal.*,(2018)	Quantitative	Asystemtodetectstress using biosensors EEG integrated with mobile development(iosdevice) usingMLtechniques	Electroencephalography(EEG)	SVM KNN Mobile Application	Not reported
Adnan*etal.*,(2015)	Quantitative	Machine Learning-based SignalProcessingUsing PhysiologicalSignalsfor StressDetection	K-nearest neighbour(KNN),and support vector machine (SVM)	Respiration, GSRHand, GSR Foot, Heart Rate andEMG	92.06% 98.41%

LimitationofML andMentalHealthStudies

LimitationsofMachineLearninginClinicalDataandSocialMedia Validation

The studies reviewed highlight several limitations ofMachine Learning (ML) when applied to clinical datasets and social media validation. One of the primary challenges is ensuring accuracy and reliability when dealing with small datasets. To enhance model performance, 10-fold cross-validation is applied, particularly when working with limited data. The iterative k-fold cross-validationtechnique isemployedtominimizeerrors,asdemonstratedby Ahuja et al. (2019).

Toreduceambiguityinpredictions, WordSenseDisambiguation(WSD) has been integrated with Support Vector Machines (SVM), improving overall accuracy (Reshma & Kinariwala, 2019). However, supervised learning methodsused forstresspredictionfromtextualdata face inherent limitations and struggle to achieve accuracy comparable to human assessment.

ChallengesinDiagnosingMentalHealthDisorderswithML

Medicalprofessionals oftenclassifydisorders basedon diagnostic uniformity, such as Major Depressive Disorder (MDD), characterized by symptoms like insomnia,confusion,depression,andfatigue.However,MLmodelsmustnow evolve to identify new subtypes of psychological conditions, including substance abuse, alcoholism, and illness trajectories (Hussain et al., 2015).

Acritical issue in ML-based medical applications is that the size ofthe dataset significantlyimpactsalgorithmicperformance. MLmodelshaveshownweaker accuracy in larger samples compared to smaller datasets. Proper implementation of cross-validation techniques helps mitigate overestimation and variability, improving the robustness of predictive models.

4. Discussion:FuturevResearchvandvRecommendation

MachineLearninginMentalStressDetectionandHealthcare

Thisstudyaimsto exploretheroleofMLalgorithms indetecting mentalstress and their specific applications in different domains. The analysis focuses on stress detection through data collected from social media posts, clinical records, blogs, student datasets, and biosignals such as EEG, ECG, and HRV. The study determines which ML algorithm performs best on particular datasets. However, the accuracyof predictions depends largely on the volume of data used for training the model. Clinical datasets, gathered over a specific period, help in generating insights and triggering early warnings for potential health abnormalities.

EffectivenessofMachineLearninginHealthApplications

AmongvariousMLtechniques, SupportVectorMachines(SVM)arewidely applied in the health sector, while Naïve Bayes is particularly effective for sentiment analysis on Facebook statuses. Comparativeanalyses indicatethat hybrid classifiers yield higher accuracy and improve overall performance compared to traditional classification methods.

ML has significantly improved healthcare outcomes by leveraging large volumesofmedicaldata. Oneofthekeychallenges inhealthcare iseffectively collecting, managing,andanalyzingthisdatafor predictionandtreatment.ML assists health professionals in decision-making by identifying patterns, fostering medical innovation, and enhancing the efficiency of clinical trials. Real-timedatafromsourceslike labtestresults,bloodpressurereadings,and family medical history, when combined with historical health records, helps develop predictive models for disease forecasting.

AccordingtotheFraminghamstudy,MLachieved 56%accuracyinpredicting long-termcardiovascular disease risks. Additionally, ML is being explored in clinical trials for drug discovery and substance abuse detection. Studies indicatethataround 90%ofdrugsfailintrials, but automateddrugdiscovery can reduce costs by nearly 70%, as reported byCarnegie Mellon University.

ML-BasedElectronicHealthRecords(EHR)andEpidemicPrediction

Machine Learning-based Electronic Health Records (EHR) enable the application of predictive models across various healthcare systems. These systems integrate structured and unstructured data, including medical images, audio recordings, text-based reports, and diagnostic scans. Advanced technologies like optical character recognition (OCR), image processing (radiology, pathology, dermatology), and natural language processing (NLP) helpconvertcomplexmedicaldataintostructuredformatsforfurtheranalysis.

AcollaborationbetweenGoogleandUniversityCollegeLondonHospital led to the development of an ML algorithm that distinguishes cancerous and healthy tissues, significantly improving radiotherapy treatments. ML is also extensively utilized for predicting and monitoring epidemic outbreaks by analyzing data from social media, web sources, and satellite imaging.

Conclusion

Ultimately, Machine Learning plays a critical role in extracting insights from complex health datasets for disease prediction, treatment recommendations, and medical analysis. However, it is essential to evaluate whether these ML-driveninsightscanbeeffectivelyinterpretedandapplied inclinicalsettings for improved healthcare outcomes.

References:

1. **Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662–679.**
2. **Chancellor,S.,&DeChoudhury,M.(2020).Methodsinpredictive techniques for mental health. *NPJ Digital Medicine, 3*(1), 1–12.**
3. **Cheng, Q.,etal. (2017).Social mediaandsuicideprevention.*Crisis,38*(4), 207–214.**
4. **Coppersmith,G.,etal.(2015).QuantifyingmentalhealthsignalsinTwitter. *CLPsych Workshop*.**
5. **CrisisTextLine.(2020).Annualreport.**
6. **DeChoudhury,M.,etal.(2013).Predictingdepressionviasocial media. *ICWSM*.**

7. **Devlin, J., et al. (2018). BERT: Pre-training for language understanding. *NAACL*.**
8. **Eskin,M.,etal.(2020).AIforsuicideriskdetection. *JMIR,22*(10),e12572.**
9. **Garcia, D., et al. (2021). COVID-19 and mental health discourse. *JMIR,23*(5), e26915.**
10. **Guntuku,S.C.,etal.(2020).MentalhealthdiscourseduringCOVID-19.*JMIR,22*(9),e22635.**
11. **Ji,S.,etal.(2023).GPT-4forsuicideriskdetection. *NatureMental Health, 1*(3), 1–9.**
12. **Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT. *arXiv:1907.11692*.**
13. **Mittelstadt,B.D.,etal.(2016).Theethicsofalgorithms. *BigData& Society, 3*(2).**
14. **Obermeyer, Z., et al. (2019). Dissecting racial bias in health algorithms. *Science, 366*(6464), 447–453.**
15. **Poria, S., etal. (2021). Multimodal sentiment analysis. *IEEE Transactions on Affective Computing, 12*(2), 1–15.**
16. **Reece,A.G.,etal.(2022).Detectingself-harminadolescentsvia ML. *JAMA Pediatrics, 176*(3), 250–257.**
17. **Rocher,L.,etal.(2019).Estimatingre-identificationrisks. *Nature Communications, 10*(1), 1–9.**
18. **Torous,J.,etal.(2021).Thefutureofmentalhealthapps. *JAMA Psychiatry, 78*(2), 115–126.**
19. **WHO.(2022).Mentalhealthatlas.**
20. **Yang,Q.,etal.(2019).Federatedlearning. *SynthesisLecturesonAI, 12*(3), 1–207.**